

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:31:59 ; Search time 41 Seconds

(without alignments)
553,556 Million cell updates/sec

Title: US-09-701-675A-3

Perfect score: 1210
Sequence: 1 MLSKGLKRRKEEKEPLA.....YLMADVLTQALRPPGGR 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 120 | 9.9 | 3164 | 1 WMBEH6 | UL36 protein - hum |
| 2 | 112.5 | 9.3 | 593 | 2 I51213 | drebrin - chicken |
| 3 | 109 | 9.0 | 688 | 3 S61249 | probable virion pr |
| 4 | 108.5 | 9.0 | 1217 | 2 T00270 | hypothetical prote |
| 5 | 107.5 | 8.9 | 440 | 2 S71795 | transcription fact |
| 6 | 107.5 | 8.9 | 708 | 1 Q08EB8 | UL80 protein - hum |
| 7 | 107 | 8.8 | 1317 | 2 T03748 | apoptosis associat |
| 8 | 106.5 | 8.8 | 512 | 2 H84310 | cobryic acid synth |
| 9 | 106.5 | 8.8 | 753 | 2 T46614 | chemotaxis protein |
| 10 | 106.5 | 8.8 | 909 | 2 S32538 | CGMP-gated cation |
| 11 | 105 | 8.7 | 607 | 2 A43776 | drebrin E2 - chick |
| 12 | 102.5 | 8.5 | 801 | 2 T29018 | phosphoprotein, sy |
| 13 | 102.5 | 8.5 | 901 | 2 A44825 | hypothetical prote |
| 14 | 102.5 | 8.5 | 908 | 2 A33480 | sarcalumenin precu |
| 15 | 101.5 | 8.4 | 676 | 1 EDBE22 | immediate-early pr |
| 16 | 101.5 | 8.4 | 1095 | 2 T00329 | hypothetical prote |
| 17 | 101.5 | 8.4 | 2774 | 2 A43359 | microtubule-associ |
| 18 | 101 | 8.3 | 433 | 2 S69999 | sterigmatocystin s |
| 19 | 101 | 8.3 | 1532 | 2 A61262 | collagen alpha 1(X |
| 20 | 100.5 | 8.3 | 676 | 1 EDBE23 | immediate-early pr |
| 21 | 100 | 8.3 | 655 | 2 S40521 | FKHR protein - hum |
| 22 | 100 | 8.3 | 1851 | 2 T19964 | hypothetical prote |
| 23 | 99.5 | 8.2 | 395 | 2 I49575 | CCAAT/enhancer bin |
| 24 | 99.5 | 8.2 | 1106 | 1 A39299 | Bassoon protein - |
| 25 | 99.5 | 8.2 | 3942 | 1 T42730 | secreted glycoprote |
| 26 | 99 | 8.2 | 498 | 1 VGBEGX | collagen, cornea-s |
| 27 | 98.5 | 8.1 | 1446 | 2 A38587 | bullous pemphigoid |
| 28 | 98.5 | 8.1 | 1433 | 2 A46053 | hypothetical prote |
| 29 | 98 | 8.1 | 337 | 2 T21055 | |

| | | | | | |
|----|------|-----|------|----------|--------------------|
| 30 | 98 | 8.1 | 1199 | 2 A40670 | nuclear envelope p |
| 31 | 98 | 8.1 | 2282 | 2 T42717 | DNA-binding protei |
| 32 | 98 | 8.1 | 3010 | 1 GNMVTV | genome polyprotein |
| 33 | 96.5 | 8.0 | 590 | 1 T35297 | probable dihydrol |
| 34 | 96.5 | 8.0 | 1207 | 2 T00378 | KIAA0641 protein |
| 35 | 96.5 | 8.0 | 1315 | 2 A56101 | collagen alpha 1(X |
| 36 | 96.5 | 8.0 | 1774 | 2 B56101 | collagen alpha 1(X |
| 37 | 96 | 7.9 | 416 | 1 SKXLAG | dermal gland prote |
| 38 | 96 | 7.9 | 477 | 2 T46304 | hypothetical prote |
| 39 | 96 | 7.9 | 936 | 2 D97630 | hypothetical prote |
| 40 | 96 | 7.9 | 1298 | 1 EDBE75 | immediate-early pr |
| 41 | 95.5 | 7.9 | 245 | 1 W4WU5 | E4 protein - human |
| 42 | 95.5 | 7.9 | 303 | 1 TVH0JD | transforming prote |
| 43 | 95.5 | 7.9 | 614 | 2 T33149 | hypothetical prote |
| 44 | 95.5 | 7.9 | 672 | 2 I40333 | tracheal colonizat |
| 45 | 95.5 | 7.9 | 1257 | 2 S28764 | neurocan precursor |

ALIGNMENTS

RESULT 1

WMBEH6
UL36 protein - human herpesvirus 1 (strain 17)

C:Species: human herpesvirus 1

C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000

C:Accession: I30085

R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; J

J. Gen. Virol. 69, 1531-1574, 1988

A:Title: The complete DNA sequence of the long unique region in the genome of herpes

A:Reference number: A30083; MUID:86274327; PMID:2839594

A:Accession: I30085

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-3164 <MGCS>

A:Cross-references: GB:X14112; NID:91944536; PIDN:CAA3231.1; PID:959536; GB:D00317

A:Gene: UL36

A:Superfamily: varicella-zoster virus gene 22 protein

| Query Match | Score | Length | DB ID | Description |
|---|-------|--------|----------|--------------------|
| 21 VDSWMDPGHAUVAQAPVASSLSFEDSLKHSLSQSEPDRLHLYVNTLRIRQAS 80 | 9.98 | 3164 | 1 WMBEH6 | UL36 protein - hum |
| 2396 ITWAVRLHADLNPTEACIAAOLPRISALTAERPLARCP---CLVAV-----DIS 2445 | 26.78 | 3164 | 2 I51213 | drebrin - chicken |
| 81 MAPAAALPPVPSPAPASV-----ADNLLASS-----DAATASNASLLE 120 | 23 | 3164 | 3 S61249 | probable virion pr |
| 2446 MTPVAVLWENPDPPGPPDVRFGSEATELTPVAGGEDVLASATDEDEPFLARALIGRPF 2505 | 23 | 3164 | 2 T00270 | hypothetical prote |
| 121 DLSHIEGSLQAPPLADEPFGSRSTIGGAAPSIGALDL-----LGPAT---GCLLDGLEG 172 | 23 | 3164 | 2 S71795 | transcription fact |
| 2506 DASLSGELFPFGHPYQORAPDQSSVNPPTGPPVDLVGARESLGSLGSLATLFLFDAPRG 2565 | 23 | 3164 | 1 Q08EB8 | UL80 protein - hum |
| 173 LFEDIDTSYDNEELNAPASEGLKPGEDPGKEAPELDEALDYLMVLTG---QALERP 231 | 23 | 3164 | 2 T03748 | apoptosis associat |
| 2566 --EPVPPRRW---ANIRHGLELASDSDSGP---APLAPPLPSTADQSPVTSQCAPRP 2616 | 23 | 3164 | 2 H84310 | cobryic acid synth |
| 232 PGP 234 | 23 | 3164 | 2 S32538 | CGMP-gated cation |
| 2617 PGP 2619 | 23 | 3164 | 2 A43776 | drebrin E2 - chick |

SEQUENCE COMPARISON

RESULT 2
drebrin - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 21-Jul-2000
R:Kojima, N.; Shitao, T.; Odate, K.
Brain Res. Mol. Brain Res. 19, 101-114, 1993

THIS PAGE BLANK (USPTO)

A:Title: Molecular cloning of a developmentally regulated brain protein, chicken drebrin
A:Reference number: 151212; MUID:93368392; PMID:8361332
A:Accession: 151213
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-593 <KOJ>
A:Cross-references: GB:S65296; NID:g410604; PIDN:AAB28012.1; PID:g410605
C:Genetics:
A:Introns: 26/3; 51/3; 100/3; 126/3; 177/2; 198/3; 218/3; 257/1; 303/1; 346/1; 536/3; 57

Query Match 9.3%; Score 112.5; DB 2; Length 593;
Best Local Similarity 25.2%; Pred. No. 1.3;
Matches 63; Conservative 18; Mismatches 76; Indels 93; Gaps 13;

QY 5 GLKRRKEEEE-----KEPLAVDSWMLDPCHAAVAQAPPAVASSSLFDLS 49
DB 329 GEORTRAPAPATPAPKSPSPSTQVAPATQOHPPFPPEDKAAEPG----- 378
50 VLKHLHSLQOSEPD-----LRHLVYVNTLRRIQASMAPAALPPVSPAPAS 98
379 -----DEPPDPAPAPATAGADVIGDLY-----TLPESEPPAPAS--EPQVETRG 423
99 VADNL--LASSDAALASMASLLEDLSHIGLSQADQPLADEGPPGRSIGAAPSLGALD 156
424 VAEPLLELMQSDGAAPAA-----TSTWLPDPAPAGP-----VPEEGT-- 462
QY 157 LIG-----PAGCLLDGDELFEDIDMSYDNEIEMA-----PASEGLKPGEDDGRKE 205
DB 463 LILDELPEPPAFPCDAEQ-----HEVEEEEEEETAGEPHPTGLGVOEGVOEGP--- 514
QY 206 EAEPLDEAEL 215
DB 515 EYPPITNGEM 524

RESULT 3

S61249
Probable v1rion protein - bovine herpesvirus 1
C:Species: bovine herpesvirus 1
C>Date: 18-Sep-1997 #sequence_revision 18-Sep-1997 #text_change 26-Aug-1999
C:Accession: S61249
R:Vicek, C.; Benes, V.; Lu, Z.; Kutish, G.F.; Paces, V.; Rock, D.; Letchworth, G.J.; Sch
submitted to the EMBL Data Library, January 1995
A:Description: Nucleotide sequence analysis of a 30-kb region of the bovine herpesvirus
A:Reference number: S61233
A:Accession: S61249
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-688 <VLC>
A:Cross-references: EMBL:248053; NID:g971311; PIDN:CAA88127.1; PID:g971328
superfamily: varicella-zoster virus gene 54 protein

Query Match 9.0%; Score 109; DB 2; Length 688;
Best Local Similarity 26.6%; Pred. No. 2.8;
Matches 45; Conservative 25; Mismatches 47; Indels 52; Gaps 7;

QY 27 DPGHAAVAQAPPAVASSSLFDLSYLKCH-----HSIQOSEPDRLHLYVNTL----- 74
DB 255 EEEGCEYLMAPVWFGDYLFESPQRLHGEVLAHALREH-----ARLCQILNTVPLKVLV 310
QY 75 -RRIOASMAPAALPPVSPAPSVADNLLASSDAALASMASLLEDLSHIGLSQAPQ 133
DB 311 GRRAEALRA-----KEPGAPSVTEKLIGEGEDAAASSAARLILTYNMKGM----- 356
QY 134 PLADEGPPGRSIGAAPSLGALDLGATGCLLDGLEGFE--DIDTS 180
DB 357 -----RHIGDISSETVS-----YIDDFAGLFDVSDVTS 386

RESULT 4

S60270
Chemical protein KIAA0596 - human (fragment)
A:Status: Homo sapiens (man)

C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00270
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A:Reference number: 214086; MUID:98290545; PMID:9628581
A:Accession: T00270
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1217 <NAG>
A:Cross-references: EMBL:AB011168; NID:g3043715; PIDN:BAA25522.1; PID:g3043716
A:Experimental source: brain
C:Genetics:
A>Note: KIAA0596

Query Match 9.0%; Score 108.5; DB 2; Length 1217;
Best Local Similarity 26.2%; Pred. No. 6;
Matches 65; Conservative 21; Mismatches 91; Indels 71; Gaps 12;

QY 22 DSWMLDPCHAAVAQAPPAVASSSLFDLSYLKHLHSLQOSEPDRLHLYVNTLRRIQASM 81
DB 881 DSSW-----APKRVATASPF--SGLQAKQSVHSIVPERH-----EAL 917
QY 82 -APA-----AALPPVSPAPASVADNLLASSDAALASMASLLEDLSHI 125
DB 918 QAPSPGALLSRIEADQDGLSLPADGPPSRPHSYQNPPTSSMAKISRST--SVGENL--- 973
QY 126 EGISQAPQ-----PLADEGPPGRS-----IGGAAPSLGALDLGPRITGCLLDGDE- 171
DB 974 -GLVAPRQAHAPIRVPSLSKLALPSRAHLVLDIPKPLPRLTAAPSPVTKGRAPGEAKR 1032
QY 172 -GLFEDIDMSYDNEIEMAPASEGLKPGP-----EDPGKEAPEDEAEIDLVMDVLVG 224
DB 1033 PGFPVGLGKAHSTTEHMACLGEGTTPKPRTECOAHGPPSPCAQQLPVSSL-----FQG 1086
QY 225 TQALERPP 232
DB 1087 PENLQPPP 1094

RESULT 5

S71795
Transcription factor CBF-2 - chicken
C:Species: Gallus gallus (chicken)
C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jul-2000
C:Accession: S71795
R:Yunasa, J.; Hirano, S.; Yamagata, M.; Noda, M.
Nature 382, 632-635, 1996
A:Title: Visual projection map specified by topographic expression of transcription f
A:Reference number: S71794; MUID:96338226; PMID:8757134
A:Accession: S71795
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-440 <YUN>
A:Cross-references: EMBL:U47276; NID:g1546783; PIDN:AAB08467.1; PID:g1546784
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F:143-234/Domain: fork head DNA-binding domain homology <FHN>

Query Match 8.9%; Score 107.5; DB 2; Length 440;
Best Local Similarity 32.3%; Pred. No. 2.1;
Matches 40; Conservative 12; Mismatches 45; Indels 27; Gaps 6;

QY 82 APAAALPPVSPAPASVADNLLASSDA-----LSASMASLLEDLSHIGLSQAP 132
DB 315 APAAALPPVSPAPAPRRAPLPPAAELATPPGYHPHGPALAAALHAAPGSGAAVAVR 374
QY 133 QPLADEGPPGRSIGAAPSLGALDLGPRATG-----CLLDG--LEGFEDIDTMYDNEIEM 187
DB 375 SPFSISIIIG--GGGPGGLGA-----GPAPGAGSCASGSGAATGSLRSLSGSL----- 421
QY 188 APAS 191
DB 422 APAA 425

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:31:59 ; Search time 41 seconds
(without alignments)
553.556 Million cell updates/sec

Title: US-09-701-675A-3
Perfect score: 1210
Sequence: 1 MSLKGLRRREEREEKEPLA.....YLMADYLVGTQALRRPPGGR 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 120 | 9.9 | 3164 | 1 WMBEH6 | UL36 protein - hum |
| 2 | 112.5 | 9.3 | 593 | 2 I51213 | drebrin - chicken |
| 3 | 109 | 9.0 | 688 | 2 S61249 | probable virion pr |
| 4 | 108.5 | 9.0 | 1217 | 2 T00270 | hypothetical prote |
| 5 | 107.5 | 8.9 | 440 | 2 S71795 | transcription fact |
| 6 | 107.5 | 8.9 | 708 | 1 Q0BBE8 | UL80 protein - hum |
| 7 | 107 | 8.8 | 1317 | 2 T03748 | apoptosis associat |
| 8 | 106.5 | 8.8 | 512 | 2 H84310 | cobrytic acid synth |
| 9 | 106.5 | 8.8 | 753 | 2 T46614 | chemotaxis protein |
| 10 | 106.5 | 8.8 | 909 | 2 S32538 | CGMP-gated calton |
| 11 | 105 | 8.7 | 607 | 2 A43776 | drebrin E2 - chick |
| 12 | 102.5 | 8.5 | 801 | 2 T29018 | hypothetical prote |
| 13 | 102.5 | 8.5 | 901 | 2 A44825 | phosphoprotein, sy |
| 14 | 102.5 | 8.5 | 908 | 2 A33280 | sarcalumenin precu |
| 15 | 101.5 | 8.4 | 676 | 1 EDBE22 | immediate-early pr |
| 16 | 101.5 | 8.4 | 1095 | 2 T00329 | hypothetical prote |
| 17 | 101.5 | 8.4 | 2774 | 2 A43359 | microtubule-associ |
| 18 | 101 | 8.3 | 433 | 2 S69999 | sterigmatocystin s |
| 19 | 101 | 8.3 | 1532 | 2 A61262 | collagen alpha 1(X |
| 20 | 100.5 | 8.3 | 676 | 1 EDBE23 | immediate-early pr |
| 21 | 100 | 8.3 | 655 | 2 S40521 | FKHR protein - hum |
| 22 | 100 | 8.3 | 1851 | 2 T19964 | hypothetical prote |
| 23 | 99.5 | 8.2 | 395 | 2 I49575 | CCAAT/enhancer bin |
| 24 | 99.5 | 8.2 | 1106 | 1 A39299 | DNA-directed DNA p |
| 25 | 99.5 | 8.2 | 3942 | 2 T42730 | Bassoon protein - |
| 26 | 99 | 8.2 | 498 | 1 VGBEGX | secreted glycoprot |
| 27 | 98.5 | 8.1 | 1446 | 2 A38587 | collagen, cornea-s |
| 28 | 98.5 | 8.1 | 1433 | 2 A46053 | bullous pemphigoid |
| 29 | 98 | 8.1 | 337 | 2 T21053 | hypothetical prote |

| | | | | | |
|----|------|-----|------|----------|--------------------|
| 30 | 98 | 8.1 | 1199 | 2 A40670 | nuclear envelope p |
| 31 | 98 | 8.1 | 2282 | 2 T42717 | DNA-binding protei |
| 32 | 98 | 8.1 | 3010 | 1 GNMATW | genome polyprotein |
| 33 | 96.5 | 8.0 | 590 | 1 T35297 | probable dihyroli |
| 34 | 96.5 | 8.0 | 1207 | 2 T00378 | KIAA0641 protein - |
| 35 | 96.5 | 8.0 | 1315 | 2 A56101 | collagen alpha 1(X |
| 36 | 96.5 | 8.0 | 1774 | 2 B56101 | collagen alpha 1(X |
| 37 | 96 | 7.9 | 416 | 1 SKXLAG | dermal gland prote |
| 38 | 96 | 7.9 | 477 | 2 T46304 | hypothetical prote |
| 39 | 96 | 7.9 | 936 | 2 D97630 | hypothetical prote |
| 40 | 96 | 7.9 | 1298 | 1 EDBE75 | immediate-early pr |
| 41 | 95.5 | 7.9 | 245 | 1 W4WLS | E4 protein - human |
| 42 | 95.5 | 7.9 | 303 | 1 TVH0UD | transforming prote |
| 43 | 95.5 | 7.9 | 614 | 2 T33149 | hypothetical prote |
| 44 | 95.5 | 7.9 | 672 | 2 I40333 | tracheal colonizat |
| 45 | 95.5 | 7.9 | 1257 | 2 S28764 | neutocan precursor |

ALIGNMENTS

RESULT 1
WMBEH6
UL36 protein - human herpesvirus 1 (strain 17)
C:Species: human herpesvirus 1
C>Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000
C:Accession: I30085
R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; P
J. Gen. Virol. 69, 1531-1574, 1988
A:Title: The complete DNA sequence of the long unique region in the genome of herpes
A:Reference number: A30083; MID:8824327; PMID:2839594
A:Accession: I30085
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-3164 <MCG>
A:Cross-references: GB:X14112; NID:q1944536; PIDN:CAA32311.1; PID:q59536; GB:D00317
C:Genetics:
A:Molecule type: DNA
C:superfamily: varicella-zoster virus gene 22 protein
Query Match 9.9%; Score 120; DB 1; Length 3164;
Best Local Similarity 26.7%; Pred. No. 2.8;
Matches 65; Conservative 23; Mismatches 107; Indels 48; Gaps 10;
QY 21 VQSWMLDPGAAVAAQAPPAVASSLSFEDLVKLKHSIQSEPDRLRLVYVNTLRITQAS 80
DB 2396 IETWAVRLHADINPIENACLAQPLRLAERPLANGPP---CLVLV-----DTS 2445
QY 81 MAPAAALPPVPSPPAASV-----ADNLLAS---DAALSMSASLIE 120
DB 2446 MPVPAVAVLWENPDPPGPPDVFVGSSEATEELPFVAGEGDVLAASATDEDFLARILGRPF 2505
QY 121 DLSHIEGLSQAPQPLADDEPPGGRSICGAAPSIGALD-----LGPAT---GCLLDGLEG 172
DB 2506 DASLISGELFPGRPPVQARPDOSPEVPMPRTGPPVDLVGAESLIGPSTLPTLFTDAPG 2565
QY 173 LPEDIDTSMYDNEIMAPASEGLKPGEDPGKREAPLEAEFLMDLVCT-QALERP 231
DB 2566 --EPVPPRW---AWIHGLEELASDSDGSP---APLAPDPPLSPADSVPTSCAPRP 2616
QY 232 PGP 234
DB 2617 PGP 2619
RESULT 2
I51213
drebrin - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 21-Jul-2000
C:Accession: I51213
R:Koijima, N.; Shitao, T.; Oyata, K.
Brain Res. Mol. Brain Res. 19, 101-114, 1993

A>Title: Molecular cloning of a developmentally regulated brain protein, chicken drebrin
A:Reference number: 151212; MUID:99368392; PMID:8361332

A:Accession: 151213

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-593 <KOU>

A:Cross-references: GB:S65296; NID:g410604; PIDN:AAB28012.1; PID:g410605

C:Genetics:

A:Introns: 26/3; 51/3; 100/3; 126/3; 177/2; 198/3; 218/3; 257/1; 303/1; 346/1; 536/3; 571/1

Query Match 9.3%; Score 112.5; DB 2; Length 593;
Best Local Similarity 25.2%; Pred. No. 1.3;
Matches 63; Conservative 18; Mismatches 76; Indels 93; Gaps 13;

OY 5 GLKKRREEEE-----KEPLAVDSMWLDPGHAAVAOAQPPVASSSLDLS 49
 ||| |
Db 329 GEQGTTRAPAEETPATPDSPSPQVAEPATREQHWPFGGEDRAAPERG----- 378
 ||| |
 50 VLKLHHSLQOSEPD-----LRLVLVNTLRRIQASMAPAALPVPSPPAAPS 98
 :||| |
Db 379 -----DEEDPDPRPAWTAGADVLGDLY-----TLPESEPSAPAAS---EPQPVETPG 423
 ||| |
OY 99 VADNL--LASSDALSAWSMAILEDLSHIEGLSQAPDLADEGPGRSGSAGASISALD 156
 ||| |
Db 424 VAEPLEIMQSDGAAPAA-----TSWPLPDTPAGFP-----VPEEETG-- 462
 ||| |
OY 157 LLG-----PATGCLDDGLEGLFEDIIDTSMYNELMA---PASEGLKRPGEGRKE 205
 ||| |
Db 463 LLAGDELPEPATPCDAEQ-----HEVEEEEEEEETAGEPHPTGLGYRGYGEGP--- 514
 ||| |
OY 206 EAPLEDAEL 215
 ||| |
Db 515 EVPIITNGEM 524

RESULT 3
S61249
probable viroion protein - bovine herpesvirus 1
C:Species: bovine herpesvirus 1
C:Date: 18-Sep-1997 #sequence_revision 18-Sep-1997 #text_change 26-Aug-1999
C:Accession: S61249
R:Vick, C.; Benes, V.; Lu, Z.; Kutlish, G.F.; Paces, V.; Rock, D.; Letchworth, G.J.; Setlow
submitted to the EMBL Data Library, January 1995
A:Description: Nucleotide sequence analysis of a 30-kb region of the bovine herpesvirus
A:Reference number: S61233
A:Accession: S61249
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-688 <VIC>
A:Cross-references: EMBL:Z48053; NID:g971311; PIDN:CAA88127.1; PID:g971328
C:Superfamily: varicella-zoster virus gene 54 protein

Query Match 9.0%; Score 109; DB 2; Length 688;
Best Local Similarity 26.6%; Pred. No. 2.8;
Matches 45; Conservative 25; Mismatches 47; Indels 52; Gaps 7;

OY 27 DPGHAAVAOAQPPVASSSLFDLSYLKLH-----HSIQOSEPDRLHLVNVLT----- 74
 ||| |
Db 255 EEGECFLTMAPVMWGGDVLFESPMQRHLGELYLAHALREH---ARICQLINTPYPLKVIV 310
 ||| |
OY 75 -RRIQASMAPAALPVPSPPAAPSVADNLIASSDAALSASMASLLEDLSHIEGLSQAPQ 133
 ||| |
Db 311 GRRAEAALRA-----KPGAPSVYEKLIGEGEDAASSANRLIKILVNMKGMM----- 356
 ||| |
OY 134 PLADEGPGRSIGGAPSLGALLDLPATGCLDDGLEGLFE--DIITS 180
 ||| |
Db 357 -----RHIGDISETVNS-----YLDRTAAGLFVDSDVPTS 386
 ||| |

RESULT 4
T00270
hypothetical protein KIAA0596 - human (fragment)
C:Species: Homo sapiens (man)

```

C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C/Accession: T00270
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,
DNA Res. 5, 31-39, 1998
A>Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A/Reference number: Z14086; MUID:98290545; PMID:9628581
A/Accession: T00270
A>Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 11217 <NAG>
A/Cross-references: EMBL:AB011168; NID:93043715; PIDN:BAA25522.1; PID:93043716
C/Genetics:
C/Note: KIA00596

Query Match          9.0%: Score 108.5; DB 2; Length 1217;
Best Local Similarity 26.2%: Pred. No. 6;
Matches 65; Conservative 21; Mismatches 91; Indels 71; Gaps 12;

QY 22 DSWMLDGGHAAVAQAAPPAVAVSSSLFEDLSVLKHLHSHLQOSEPDLRHLVLYVNTLRRIQASM 81
Db 881 DSSW-----APKRVATAAPSF--SGLQKQSVSHVLPQENH-----EASL 917

QY 82 -APA-----AALPVPSPPAPASVADNLLASDAASMASASMASLIEDLSHI 125
Db 918 QAAPSGALLSREIEBAQDGLSLPPADGPSPSRPHSYQNPPTSSMAKISRST-SVGENL--- 973

QY 126 EGLSQAQP-----PLAEGPPGRS-----IGAPASLGALDLDLPATGCLDDLE- 171
Db 974 -GIVAEQAAPAHPIVPSLSKIALPSRAHLYVIDIPKPIBDRPTLAAFSPIYKGRAPGAER 1032

QY 172 -GLFEDIDTSMYDNELMAPSEGLKPGP-----EDGPGKEBAPELDEALDYLMDVYG 224
Db 1033 PGFVGVGAKAHSTIERMACLGEGTTPKPRTECQAHPSPSSCAQQLPVSSD-----FOG 1086

QY 225 TQALERP 232
Db 1087 PENIQAPP 1094

RESULT 5
S71795
transcription factor CBF-2 - chicken
C/Species: Gallus gallus (chicken)
C/Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jul-2000
C/Accession: S71795
R:Yuasa, J.; Hiranano, S.; Yamagata, M.; Noda, M.
Nature 382, 632-635, 1996
A>Title: Visual projection map specified by topographic expression of transcription f
A/Reference number: S71794; MUID:96338226; PMID:8757134
A/Accession: S71795
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1440 <YUAS>
A/Cross-references: EMBL:U047276; NID:91546783; PIDN:AAB08467.1; PID:91546784
C/Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F/143-234/Domain: fork head DNA-binding domain homology <FHD>

Query Match          8.9%: Score 107.5; DB 2; Length 440;
Best Local Similarity 32.3%: Pred. No. 2.1;
Matches 40; Conservative 12; Mismatches 45; Indels 27; Gaps 6;

QY 82 APAALAPPVSPAPASVADNLLASDA-----LSASMASLIEDLSHIEGLSQAP 132
Db 315 APAALAPPPPPPPRRRAPLPAALAKTPGYPPLGPPALAAASHAAKPGSGAAVAR 374

QY 133 QPIADECPGPRSIGAAPSLGALDILGPATG---CLLDG-LEGLFEDIDTSMYDNELW 187
Db 375 SPFSIEIITG---GGPGLGA---GPAPGAGSCASQSGAATGLSRISLGSGL----- 421

QY 188 APAS 191
Db 422 APAA 425

```

RESULT 6

Q00BB8

U180 protein - human cytomegalovirus (strain AD169)

N:Contins: capsid assembly protein; viral proteinase (EC 3.4.21.-)

C:Species: human cytomegalovirus, human herpesvirus 5

A:Note: host Homo sapiens (man)

C:Date: 31-Dec-1990 #sequence, revision 31-Dec-1990 #text_change 16-Jul-1999

C:Accession: S09843; S51034; S51035

R:Chen, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerry, R.; Hornsneil, T.;

M.; Barrell, B.G.

Curr. Top. Microbiol. Immunol. 154, 125-169, 1990

A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A:Reference number: S09749; M0ID:90269039; PMID:2161319

A:Accession: S09843

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-708 <CHE>

A:Cross-references: EMBL:X17403; NID:g59591; PIDN:CAA35353.1; PID:g1780857

A:Note: possible protein-coding frames are given

A:Note: the DNA sequence was submitted to EMBL, December 1989, in computer-readable form

A:Stevens, J.T.; Mappelli, C.; Tsao, J.; Hall, M.; O'Boyle II, D.; Weinheimer, S.P.; Dilla

R.; J. Biochem. 226, 361-367, 1994

A:Title: In vitro proteolytic activity and active-site identification of the human cytom

A:Reference number: S51034; M0ID:95094793; PMID:8001353

A:Accession: S51034

A:Molecule type: protein

A:Residues: 110-131, 'X', 133-134 <ST2>

A:Note: assembly protein release site (Ala-256-Ser-257) and maturation site (Ala-643-Ser

C:Superfamily: cytomegalovirus capsid assembly protein

C:Keywords: capsid assembly; hydrolase; serine proteinase

F:1-236/Product: viral proteinase #status predicted <PMAT>

F:336-708/Product: capsid assembly #status predicted <CAP>

F:132/Active site: Ser #status experimental

F:256-257/Cleavage site: Ala-Ser (viral proteinase) #status experimental

F:643-644/Cleavage site: Ala-Ser (viral proteinase) #status experimental

Query Match

Best Local Similarity 27.1%; Pred. No. 3.7;

Matches 69; Conservative 24; Mismatches 85; Indels 77; Gaps 14;

QY 25 WIDPGRHAAV-AQAPPAVASS-SLPDLSVLKLSHQSEDLRH---LVLYVNTLRIOA 79

Db 199 WQRCSTAVDASGDPKPSDYGILGNSVALY--IRERLPKLYDKQVGVTERESYVKA 256

QY 80 SMAPAAAL-----PPY--PSP-----PAAPSV 99

Db 257 SVSPBAACDIKKAASERSGDSRSQAATPAAGARVPSSSPVPVPPSPVOPALPAPSPV 316

QY 100 ADNLASDAALASMASLLEDSHTEGLSQAPQPL-----ADEGPPGRSIGGAAPSLGA 154

Db 317 --LPAESPSPSLSPSEPAEAMSMH-----PLSAVPAATAPGATVAGASPAVSS 364

QY 155 LDLGATATGCLDD-----GLEGLFEDIDTSMYDNEIMAPASEGLAKGP-EDGCKEAP 209

Db 365 --LAWPHDGYTLPEKAFSLGASRSAPVMPYGAVAAPASAPADLPPLSPYASGAPV 422

QY 210 LDEAEL-----DYL 218

Db 423 VGYDQILAAHFADIV 437

RESULT 7

T03748

apoptosis associated tyrosine kinase - mouse

C:Species: Mus musculus (house mouse)

C:Date: 24-Mar-1999 #sequence, revision 24-Mar-1999 #text_change 08-Oct-1999

C:Accession: T03748

R:Goazza, E.; Baker, S.J.; Vora, R.K.; Reddy, E.P.

Oncogene 15, 3127-3135, 1997

A:Title: AATK: A novel tyrosine kinase induced during growth arrest and apoptosis of my

A:Reference number: Z15052; M0ID:98105706; PMID:9444961

A:Accession: T03748

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1317 <GAO>

A:Cross-references: EMBL:AF011908; NID:g2459992; PIDN:AA71837.1; PID:g2459993

C:Genetics:

A:Gene: AATK

Query Match 8.8%; Score 107; DB 2; Length 1317;

Best Local Similarity 23.7%; Pred. No. 8.4;

Matches 58; Conservative 20; Mismatches 79; Indels 88; Gaps 10;

QY 78 QASMAPAALPPVSP-----PAAPSVADN-----LNASDA 109

Db 746 EGSAEPOLPLPSVPSPSCGASLPSEBASADILPASPAPAGSWTVPEPAPTLSSGS 805

QY 110 ALSMASLLEDSHTEGLSQAPQPLADEPPGRSIG-----GAASGLADL 157

Db 806 SLQGEAPS-SEDDTTEATSGVFTDLSSDGPHTKEGIVPALRSLOKQVTPDLSLDI 864

QY 158 LGPAT--GC-----LLDGL-----EGLPEDIDTSMYDNEIMAPAS 191

Db 865 PSSASDGGCEVLSPSAAGPPGCGPRAVDSCYDTEENESPEFVLKEAHESSEPPAFGPAS 924

QY 192 EGIKPGED-----GPKKEADELDEA---ELDYIMDYLVG-----TGALER 230

Db 925 EGSEPGDPLSVSLGLSKSPYRDSAPYPSDASEPFGPEKHSIGDSQKEDLRS 984

QY 231 PPGG 235

Db 985 PPSPG 989

RESULT 8

H84310

cobyrlic acid synthase [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C:Date: 02-Feb-2001 #sequence, revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: H84310

R:Ng, W.V.; Kennedy, S.P.; Mahafraz, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky

Jung, K.H.; Alam, M.; Freltas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;

A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; M0ID:20504483; PMID:11016950

A:Accession: H84310

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-512 <STO>

A:Cross-references: GB:AE004437; NID:g10581058; PIDN:AA619852.1; GSPDB:GN00138

C:Genetics:

A:Gene: cblp

C:Superfamily: probable cobyrlic acid synthase

Query Match 8.8%; Score 106.5; DB 2; Length 512;

Best Local Similarity 24.5%; Pred. No. 3;

Matches 70; Conservative 35; Mismatches 84; Indels 97; Gaps 17;

QY 27 DPGHAAVQAAPPAVASSSLPDLVLKLSHQSEDLRLHLYVNTLRIO--ASMAPA 84

Db 244 DPG-----LPAEDSVSLPDPSE-RVVDGCGDGVDPDASVYAVPHLPHISNFTDLAPL 295

QY 85 AALP-----PVSP-----PAAPSVADNLASSDAL----- 111

Db 296 ARTPGVAVAYQPLADAPLAADAVVLPCTKTVVDLPAARDAGLGRALRAFDPGIVGLCG 355

QY 112 -----SASMASLLEDSHTEGLSQAP-----QP-----LADEGPPGRSIG 147

Db 356 YQMLGDRIRIDATFESPADDLGAVGVLPVETAPDPDRVEATRELADCG-----ALGG 411

QY 148 AAPSLCALDL-----LGPAT-----GCLLDGLEGIF-EDIDTSMYDNEIMAP 189

DB 412 ATGANTGTEIHMGRITVDGVPQVGPASARGRLGTYLHGLFENGDAARRGFRADVAA 471
 QY 190 ASEGK-KPGEDGPGKEAPELDEALDYLDVLTGQALRRPGR 234
 DB 472 A--GVEOPAPADTP--DQSPS-DAAAA--LVDADHVDLPLGVPPAP 510

RESULT 9

chemotaxis protein chea [imported] - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa

C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 31-Dec-2000

C:Accession: T46614; F83463

R:Kato, J.; Nakamura, T.; Kuroda, A.; Ohtake, H.

submitted to the EMBL Data Library, April 1998

A:Description: Cloning, sequence and characterization of chemotaxis genes in Pseudomonas

A:Reference number: 223079

Accession: T46614

Status: preliminary; translated from GB/EMBL/DBJ

Molecule type: DNA

A:Residues: 1-753 <KAT>

A:Cross-references: EMBL:AB012767; PIDN:BAA33549.1

A:Experimental source: strain PA01

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bz

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; Lim,

; Lorry, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: F83463

A>Status: preliminary

A:Residues: 1-753 <STO>

A:Cross-references: GB:AE004575; GB:AE004091; NID:99947404; PIDN:AAG04847.1; GSPDB:GNO01

A:Experimental source: strain PA01

C:Genetics:

A:Gene: chea; PA1458

C:Superfamily: chemotaxis protein chea

Query Match 8.8%; Score 106.5; DB 2; Length 753;

Best Local Similarity 25.7%; Pred. No. 4.7;

Matches 59; Conservative 29; Mismatches 101; Indels 41; Gaps 10;

DB 2 LSKGKRRKEE--EKEPLAVDSWMLDPGH-AAVAQAPPAVASSSLDLSTLKHHSLO 58

DB 119 LLAALARIAPPEGAPAPVQAPPAVPAEPAPPAQASSDITDDEFEOQLDAIQ 178

DB 59 QSEPDRLHLVVTLRIRIQASMAPAALPPVSPAPAPSVADNLLASSDAALSASMASL 118

DB 179 GDE-----APASAVAPAPAPADEISD-----AEFAL 207

QY 119 LEDLSHIEGLSQAPPLADEGP-PGRSIGAAPSLGALDILGPATCLLD--GLEGLFE 175

DB 208 LDQL-HGKGFVPPAPVSAEPAQVPAEVAEPAAAAAGD-DISDEFALLDELHG-KGKFG 264

QY 176 DIDSMYVNEIAPASEGLKRGPE-DGPGKEAPLDEALDYLDVLTG 224

DB 265 DVPEAGTPAAPAAAAAPAAPEQKAPAAAGDEISDDEFESLDELHG 314

RESULT 10

CGMP-gated cation channel 2, rod - human

C:Species: Homo sapiens (man)

C:Date: 22-Nov-1993 #sequence_revision 17-Nov-1995 #text_change 16-Jul-1999

C:Accession: S32538

R:Chen, T.Y.; Peng, Y.W.; Dhallan, R.S.; Ahmed, B.; Reed, R.R.; Yau, K.W.

Nature 362, 764-767, 1993

A:Title: A new subunit of the cyclic nucleotide-gated cation channel in retinal rods.

A:Reference number: S32538; MUID:93226050; PMID:7682292

A:Accession: S32538

A>Status: preliminary

A:Molecule type: mRNA

Query Match 8.8%; Score 106.5; DB 2; Length 909;

Best Local Similarity 21.7%; Pred. No. 5.9;

Matches 51; Conservative 34; Mismatches 89; Indels 61; Gaps 10;

A:Residues: 1-909 <CHE>

C:Superfamily: CAMP receptor protein cyclic nucleotide-binding domain homology

C:Keywords: CGMP binding

F:620-742/Domain: CAMP receptor protein cyclic nucleotide-binding domain homology <CA

Query Match 8.8%; Score 106.5; DB 2; Length 909;

Best Local Similarity 21.7%; Pred. No. 5.9;

Matches 51; Conservative 34; Mismatches 89; Indels 61; Gaps 10;

DB 11 EEEKEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVLKLSHQSEPDRLHLVLT 70

DB 88 EEEAKEP---QDW-----AETKEPEAEFAASSGVPARKQHEVVEDT----- 131

QY 71 VNTLRRIQASMAPAALPPVPS--PPAPSVADNLLASSDAALS--ASMASLEDLSHI 125

DB 132 -----ADSCPLAAEENPSTVLPSPSPAKSDTLIYSSASGTHRRKLPSEDEAEEL 183

QY 126 EGLSQAPPLADEGP-----GRTSGAAPSLGALDILGPATCCLLDLGLF----- 174

DB 184 KALSPAESPVVAMSDPTTPKDTDGDRASATAS-----TNSATINDRLQELVLTFK 234

QY 175 --EDIDTSMYDNEIWM-----APASEGLKPGEDGPGKEAPELDEALDYLT 218

DB 235 ERTEKVRKRLIDPVTSDSESPKSPAKKAPAPADTPAPAE-PVEEHYCDML 288

RESULT 11

A43776

drebrin B2 - chicken

C:Species: Gallus gallus (chicken)

C:Date: 01-Dec-1993 #sequence_revision 30-Jan-1993 #text_change 01-Dec-2000

C:Accession: A43776; B43776; I51212; I50221

R:Kojima, N.; Kato, Y.; Shitro, T.; Oyata, K.

Brain Res. Mol. Brain Res. 4, 207-215, 1988

A:Title: Nucleotide sequences of two embryonic drebrins, developmentally regulated br

A:Reference number: A43776

A:Accession: A43776

A:Molecule type: mRNA

A:Residues: 1-607 <KOJ>

A:Cross-references: GB:M36961; NID:9211725; PID:9211726

A:Accession: B43776

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-315,359-607 <KO2>

R:Kojima, N.; Shitro, T.; Oyata, K.

Brain Res. Mol. Brain Res. 19, 101-114, 1993

A:Title: Molecular cloning of a developmentally regulated brain protein, chicken dreb

A:Reference number: I51212; MUID:93368392; PMID:8361332

A:Accession: I51212

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-132 <KO3>

A:Cross-references: GB:S65230; NID:9410591; PIDN:AAB28010.1; PID:9410592

Query Match 8.7%; Score 105; DB 2; Length 607;

Best Local Similarity 24.7%; Pred. No. 4.7;

Matches 62; Conservative 18; Mismatches 77; Indels 94; Gaps 13;

DB 5 GLKRRKEEER-----KEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLS 49

DB 342 GEQRTAPAEETPATPKDSPPSYVAEPATTEQHWPPGEGDKAAEPG----- 391

QY 50 VLKLSHQSEPD-----LRHLVVTLRRIQASMAPAALPPVSPAPAS 98

DB 392 -----DEPDPPRPAPWATAGAVLGLV-----TLSPSPAPAPAS--EPQVETPG 436

QY 99 VADNL--LASSDAALSMALEDLSIEBSQAPPLADEGPGRSIGGAASGLALD 156

DB 437 VAEPLIELMQSDGAAPAA-----TSTWLPDTPAGP-----VPEEGT-- 475

QY 157 LIG-----PATCLDLDGLGLFEDIDTSMYDNEIWA-----PASGLKPGEDGPGK 204

DB 476 LIGDELPEPPATTCDAEQ-----HEVEEEEEEEETATGEPHTGLGYGEGYQEGP-- 528

R;Lafer, E.; Zhou, S.; Sousa, R.; Tannery, N.H.

1235H1EGUSYAFQPLKAD-EGFPGRSLGGAAPSDGAL--DLIGPATCGCLLDGLEGLEEDIDVF 1/9

Db 133 GDEGSLQEEERQELSSGEGPGEEAAGLGLPSEGAASGEMQGAGGKVPKEAEGVLGDSFV 192

OY 180 SMYDNELMAPASEGLKPGPED-----GPGK-EEAPELDEAELDYLMVYVGTQALERP 231
DB 193 QGAAAEFTAPEASGIAPSSEDEQIHTLEEGEGKSGSPGPDHGTPELDGTPD--GASAGEEP 249

RESULT 15

EDBE22

Immediate-early protein IER2.9 - bovine herpesvirus 1 (strain K22)

N:Alternate names: early protein ER2.6; p135 protein

C:Species: bovine herpesvirus 1

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 17-Mar-2000

C:Accession: A38209

R:Wirth, U.V.; Fraefel, C.; Vogt, B.; VLcek, C.; Paces, V.; Schwyzler, M.

J. Virol. 66, 2763-2772, 1992

A:Title: Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1 are 3' coterm

A:Reference number: A38209; MUID:92219360; PMID:1313901

A:Accession: A38209

Molecule type: DNA

Accessions: 1-676 <MIR>

Cross-references: GB:M84464; NID:9330767; PIDN:AAA6061.1; PID:9330768

C:Superfamily: herpesvirus immediate-early protein IE110; RING finger homology

C:Keywords: DNA binding; immediate-early protein; transcription regulation; zinc finger

F:9-57/Domain: RING finger homology <RING>

F:13-51/Region: zinc finger C3HC4 motif

F:284-331/Region: acidic

Query Match

8.4%; Score 101.5; DB 1; Length 676;

Best Local Similarity 25.2%; Pred. No. 9.4;

Matches 63; Conservative 26; Mismatches 86; Indels 75; Gaps 11;

OY 7 KKKREEEKEPELAVDSWMLDPCGHAVAQAPPAVASSLFDLSVLEKHHSLQOSEPDLRH 66
DB 462 RRRRTTELDRAPT-----PAPAPAPADISTWIDL----- 493
OY 67 LVLVNTLRRIQASMAPAALPP-----VPSPAPASVADNLLASSDAL 111
DB 494 ----ANAPAR-PADPAPAAAGPALAGAOIGTPAAAAAVTAAAPASVARG--SAPSPAV 546
OY 112 SASMASLLEDLSHIGLSQAPOLADEGPPGRSISGAAPSLGALDLPATGCLDDGLE 171
DB 547 TAAATGTAAAIIS----TRAPTP---SPAGRAPAADPRRAGAPALAGAAAEAGRGNP 597
OY 172 GLFEDIDSMYDNELMAPASEGLKPGPEDGPGKEAPELDEAELDYLMVYVGT-----Q 226
DB 598 GERRPASAM-----ARGDLDPGPSSSAQKRRRTTEYVAA--WRESLIGTPRRSSA 647
OY 227 ALERPPGPGR 236
DB 648 ALAPQPG-CR 656

Search completed: August 19, 2003, 09:32:51

Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:31:59 ; Search time 85 Seconds

(without alignments)
440.700 Million cell updates/sec

Title: US-09-701-675A-3

Perfect score: 1210

Sequence: 1 MLSKLRKREERKEPLA.....YLMVVGTOALRPPEGR 236

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

```
1: A_Geneseq_19Jun03:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-----------------------------|
| 1 | 1210 | 100.0 | 236 | 21 | AAV44362 Human cell cycle r |
| 2 | 1210 | 100.0 | 278 | 22 | AAW25550 Human protein sequ |
| 3 | 1206 | 99.7 | 236 | 22 | AAW93724 Human polypeptide, |
| 4 | 1206 | 99.7 | 236 | 22 | AAW89292 Human secreted pro |
| 5 | 1203 | 99.4 | 236 | 20 | AAV31829 Human adult blood |
| 6 | 1198 | 99.0 | 236 | 20 | AAV02619 Amino acid sequenc |
| 7 | 1164 | 96.2 | 236 | 20 | AAV36004 Extended human sec |
| 8 | 335.5 | 27.7 | 116 | 22 | ABG14355 Novel human diagno |
| 9 | 289 | 23.9 | 58 | 20 | AAV12453 Human 5' EST secre |

| | | | | | |
|----|-------|------|------|----|-----------------------------|
| 10 | 213 | 17.6 | 196 | 22 | AAW93922 Human polypeptide, |
| 11 | 213 | 17.6 | 196 | 22 | AAW35402 Replication protei |
| 12 | 208.5 | 17.2 | 314 | 22 | ABW65076 Hypoxia-induced pr |
| 13 | 170 | 14.0 | 241 | 22 | AAW39725 Human polypeptide |
| 14 | 170 | 14.0 | 254 | 22 | AAW41511 Human polypeptide |
| 15 | 168 | 13.9 | 254 | 21 | AAW58258 Lung cancer associ |
| 16 | 168 | 13.9 | 254 | 22 | AAW73682 Human colon cancer |
| 17 | 165 | 13.6 | 241 | 22 | AAW2898 Human haematopoiet |
| 18 | 165 | 13.6 | 241 | 22 | AAW2898 Human haematopoiet |
| 19 | 163.5 | 13.5 | 157 | 22 | AAW74742 Human CD4 protein |
| 20 | 152 | 12.6 | 237 | 23 | AAW2897 Human colon cancer |
| 21 | 129.5 | 10.7 | 322 | 22 | ABW22868 Mouse haematopoiet |
| 22 | 116 | 9.6 | 644 | 22 | ABW3698 Novel human diagno |
| 23 | 116 | 9.6 | 789 | 22 | ABW3698 Novel human diagno |
| 24 | 115 | 9.5 | 255 | 22 | AAU18154 Human prostate can |
| 25 | 115 | 9.5 | 255 | 22 | AAU18154 Human prostate can |
| 26 | 114 | 9.4 | 42 | 23 | ABW92575 Human DNA-binding |
| 27 | 114 | 9.4 | 42 | 23 | ABW92575 Human DNA-binding |
| 28 | 114 | 9.4 | 2038 | 23 | AAW25098 Human kinase and p |
| 29 | 114 | 9.4 | 2181 | 22 | AAW78959 Human protein SEQ |
| 30 | 114 | 9.4 | 2429 | 22 | AAW79943 Human protein SEQ |
| 31 | 114 | 9.4 | 2523 | 22 | AAW21713 Human PKIN-8 prote |
| 32 | 114 | 9.4 | 2545 | 23 | ABW98406 Human protein kina |
| 33 | 113 | 9.3 | 983 | 20 | AAW09513 Human NOV5, MAST20 |
| 34 | 112 | 9.3 | 791 | 23 | ABW5089 Mouse JMT protein |
| 35 | 112 | 9.3 | 802 | 22 | ABW28366 Hypoxia-repressed |
| 36 | 108.5 | 9.0 | 248 | 22 | ABW04071 Novel human diagno |
| 37 | 108.5 | 9.0 | 1101 | 22 | ABW04071 Novel human diagno |
| 38 | 108.5 | 9.0 | 1219 | 22 | AAW50136 Human GTPase activ |
| 39 | 108.5 | 9.0 | 1279 | 22 | AAW78969 Novel human diagno |
| 40 | 108.5 | 9.0 | 1464 | 22 | AAW41039 Human protein SEQ |
| 41 | 107.5 | 8.9 | 1444 | 22 | ABW15667 Human polypeptide |
| 42 | 106.5 | 8.8 | 663 | 18 | AAW29150 Novel human diagno |
| 43 | 105 | 8.7 | 175 | 22 | ABW15339 Dual-specific murt |
| 44 | 104 | 8.6 | 113 | 23 | AAW22900 Novel human diagno |
| 45 | 104 | 8.6 | 443 | 23 | ABW40426 N-terminal region |

ALIGNMENTS

| | | |
|----------|---|---|
| RESULT 1 | AAV44362 | standard; protein; 236 AA. |
| ID | AAV44362 | |
| XX | AAV44362 | |
| AC | AAV44362 | |
| XX | 14-MAR-2000 | (first entry) |
| DT | 14-MAR-2000 | (first entry) |
| XX | Human cell cycle regulation protein-3. | |
| XX | Human cell cycle regulation protein-3. | |
| DE | Human cell cycle regulation protein-3. | |
| XX | CECRP-3; cell cycle regulation protein-3; cell proliferation; | |
| KW | cell proliferative disease; cancer; atherosclerosis; cirrhosis; | |
| KW | hepatitis; psoriasis; immune system disorder; allergy; asthma; | |
| KW | acquired immune deficiency syndrome; Crohn's disease; Blast method; | |
| KW | rheumatoid arthritis; gene therapy; chromosomal mapping. | |
| XX | Homo sapiens. | |
| OS | Homo sapiens. | |
| XX | Homo sapiens. | |
| XX | Homo sapiens. | |
| FT | Key | Location/Qualifiers |
| FT | Modified-site | 44 |
| FT | Modified-site | /note= "Potential phosphorylation site" |
| FT | Modified-site | 60 |
| FT | Modified-site | /note= "Potential phosphorylation site" |
| FT | Modified-site | 73 |
| FT | Modified-site | /note= "Potential phosphorylation site" |
| FT | Modified-site | 98 |
| FT | Modified-site | /note= "Potential phosphorylation site" |
| FT | Modified-site | 117 |
| FT | Modified-site | /note= "Potential phosphorylation site" |
| FT | Modified-site | 123 |
| FT | Modified-site | /note= "Potential phosphorylation site" |
| FT | Modified-site | 180 |

FT /note= "Potential phosphorylation site"
 XX
 XX W0964593-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99MO-US12906.
 XX
 PR 08-JUN-1999; 98US-0088695.
 XX
 PA (INCYTE) INCYTE PHARM INC.
 PI Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR,
 PI Patterson C;
 DR WPI; 2000-105687/09.
 DR N-PSDB; AA229482.
 XX
 PT Novel regulatory proteins, for diagnosis, treatment and prevention of
 PT cell proliferative and immune system diseases
 XX
 PS Claim 1; Page 70; 88pp; English.
 XX
 CC The present sequence is cell cycle regulation protein-3 (CECRP-3). Prints
 CC analytical method was used to identify this protein. CECRPs are
 CC activators of cell proliferation or inhibitors of cellular processes that
 CC modulate proliferation. They are used to treat or prevent cell
 CC proliferative diseases like cancers, atherosclerosis, cirrhosis,
 CC hepatitis, psoriasis, immune system disorders (e.g. acquired immune
 CC deficiency syndrome, allergy, asthma, Crohn's disease,
 CC rheumatoid arthritis). Antibodies are raised to screen for specific
 CC binding agents. The corresponding nucleic acid is used in gene therapy,
 CC chromosomal mapping and isolation of related sequences.
 CC
 SQ Sequence 236 AA;
 Query Match 100.0%; Score 1210; DB 21; Length 236;
 Best Local Similarity 100.0%; Pred. No. 2.4e-93;
 Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLSGKLRKREERKEEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVLKIHSLQDS 60
 DB 1 MLSGKLRKREERKEEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVLKIHSLQDS 60
 QY 61 EPDRLHVLVYVNTLRRIQASMAAPAAAPPPVSPPAASVADNLASSDAALSASMASLLE 120
 DB 61 EPDRLHVLVYVNTLRRIQASMAAPAAAPPPVSPPAASVADNLASSDAALSASMASLLE 120
 QY 121 DLSHIEGSQLAPQPLADGPGPGRSIGGAAPSLGALDLGPRATGCLLDGLEGLEFEDIDTS 180
 DB 121 DLSHIEGSQLAPQPLADGPGPGRSIGGAAPSLGALDLGPRATGCLLDGLEGLEFEDIDTS 180
 QY 181 MYDNLMPASSEGKLPKPEDEPGKEAPELDEAELIDYLMVLYGQALERPGR 236
 DB 181 MYDNLMPASSEGKLPKPEDEPGKEAPELDEAELIDYLMVLYGQALERPGR 236
 RESULT 2
 AAM25550
 ID AAM25550 standard; Protein: 278 AA.
 XX
 AC AAM25550;
 XX
 DE Human protein sequence SEQ ID NO:1065.
 XX
 DT 16-OCT-2001 (first entry)
 XX
 XX Human: cancer; ulcer; HIV infection; human immunodeficiency virus;
 KM antiinflammatory; antirheumatic; antiallergic; immunosuppressive;
 KM antibacterial; endocrine; cardiac; central nervous system; vitruide;
 KM anti-HIV; fungicide; antimitagen; cardiovascular; antineoplastic; anaemia;
 KM antiagregant; haemostatic; vulnery; antiallergic; osteopathic; eczema;
 KM dermatological; antiallergic; antiaesthetic; antidiabetic; cytostatic;

KM neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KM immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KM antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KM cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KM genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KM thrombocytopenia; osteoporosis; severe combined immunodeficiency;
 KM allergic rhinitis; diabetes; multiple sclerosis; depression;
 KM Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KM neurological disorder.
 OS Homo sapiens.
 XX
 XX W0200153455-A2.
 XX
 PD 26-JUL-2001.
 XX
 PF 22-DEC-2000; 2000MO-US35017.
 XX
 PR 23-DEC-1999; 99US-0471275.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 PI
 DR WPI; 2001-457603/49.
 DR N-PSDB; AAH99491.
 XX
 PT Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection
 XX
 XX Claim 20; Page 214; 1217pp; English.
 XX
 CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiac;
 CC central nervous system; vitruide; anti-HIV; fungicide; antimitagen;
 CC cardiovascular; antineoplastic; antiagregant; haemostatic; vulnery;
 CC anticancer; osteopathic; dermatological; antiallergic; antiaesthetic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 CC
 SQ Sequence 278 AA;
 Query Match 100.0%; Score 1210; DB 22; Length 278;
 Best Local Similarity 100.0%; Pred. No. 3e-93;
 Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLSGKLRKREERKEEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVLKIHSLQDS 60
 DB 43 MLSGKLRKREERKEEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVLKIHSLQDS 102
 QY 61 EPDRLHVLVYVNTLRRIQASMAAPAAAPPPVSPPAASVADNLASSDAALSASMASLLE 120
 DB 103 EPDRLHVLVYVNTLRRIQASMAAPAAAPPPVSPPAASVADNLASSDAALSASMASLLE 162
 QY 121 DLSHIEGSQLAPQPLADGPGPGRSIGGAAPSLGALDLGPRATGCLLDGLEGLEFEDIDTS 180
 DB 163 DLSHIEGSQLAPQPLADGPGPGRSIGGAAPSLGALDLGPRATGCLLDGLEGLEFEDIDTS 222

QY 181 MYDNEIMAPASEGLKPGEDGPKGEAPELDEAELDYIMDVLTGTOALERPGR 236
DB 223 MYDNEIMAPASEGLKPGEDGPKGEAPELDEAELDYIMDVLTGTOALERPGR 278

RESULT 3
AAM93724

ID AAM93724 standard; Protein: 236 AA.

AC AAM93724;

DT 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 3677.

KM Human: full length cDNA; cDNA synthesis; oligo-capping.

OS Homo sapiens.

PN EP130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

PA (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR N-PSDB; AAK94674.

PS Claim 8; SEQ ID NO 3677; 1380pp + sequence listing; English.

CC The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a polypeptide encoded by a full length human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

QY 181 MYDNEIMAPASEGLKPGEDGPKGEAPELDEAELDYIMDVLTGTOALERPGR 236
DB 181 MYDNEIMAPASEGLKPGEDGPKGEAPELDEAELDYIMDVLTGTOALERPGR 236

RESULT 4
AAG89292

ID AAG89292 standard; Protein: 236 AA.

AC AAG89292;

DT 11-SEP-2001 (first entry)

DE Human secreted protein, SEQ ID NO: 412.

KM Human: secreted protein; gene therapy; vaccine; treatment; diagnosis; GENSET.

OS Homo sapiens.

PN MO200142451-A2.

PD 14-JUN-2001.

PF 07-DEC-2000; 2000MO-IB01938.

PR 08-DEC-1999; 99US-0169629.

PR 06-MAR-2000; 2000US-0187470.

PA (GENSET) GENSET.

PI Dumas Milne Edwards J, Bougueleret L, Jobert S;

PI WPI: 2001-367870/38.

DR N-PSDB; AAH64895.

PS Full length GENSET human nucleic acids encoding potentially secreted proteins, useful in gene therapy and vaccination against a variety of diseases, and for diagnosis of those diseases -

PS Claim 21; Page 889-890; 921pp; English.

CC The invention relates to full length GENSET human nucleic acids encoding potentially secreted proteins. The nucleic acids and the polypeptides they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate GENSET gene expression. For example, they be used to treat disorders associated with decreased GENSET gene expression by rectifying mutations or deletions in a patient's genome that affect the activity of GENSET or by supplementing the patient's own production of GENSET polypeptides. Conversely, antisense nucleic acid molecules may be administered to down regulate GENSET expression by binding with the cells' own genes and preventing their expression. The sense and antisense nucleic acids may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples, and hence to determine which patients may be in need of restorative therapy.

CC The GENSET polypeptides may be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of GENSET polypeptide expression and activity. The present sequence is a GENSET polypeptide of the invention.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

```

DB      61 EEDLRHLVLYVNTLRRIQASMAPPAALPPVSPPAAPSVADNLLASSDAALSASMASLIE 120
QY      121 DLSHIEGLSQAPQPLADEGPPGRSISGAAPSLGALDLGPATGCLLDGLEGLEFEDIDTS 180
DB      121 DLSHIEGLSQAPQPLADEGPPGRSISGAAPSLGALDLGPATGCLLDGLEGLEFEDIDTS 180
QY      181 MYDNEMLAPASBGLKPGPEDGPKKEAPLDEALDEILDYIMDVLYGTQALERPPGGR 236
DB      181 MYDNEMLAPASBGLKPGPEDGPKKEAPLDEALDEILDYIMDVLYGTQALERPPGGR 236

RESULT 5
AA31829
ID      AAY31829 standard; Protein; 236 AA.
XX
AC      AAY31829;
XX
XX      06-DEC-1999 (first entry)
DT
XX
DE      Human adult blood secreted protein g21_1.
XX
XX      Secreted protein; g21_1; human; therapy; diagnosis; vaccine; blood.
XX
OS      Homo sapiens.
XX
PH      Key
XX      Location/Qualifiers
XX      19..21
XX      /note= "signal peptide"
XX      67..79
XX      /note= "alternative signal peptide"
XX      32..236
XX      /note= "mature protein"
XX      80..236
XX      /note= "alternative mature protein"
XX      1
XX      /note= "encoded by AWG"
XX      Misc-difference 137
XX      /note= "encoded by GAS"
XX      40
XX      /note= "a putative transmembrane domain is centered
XX      around this residue"
XX      80
XX      /note= "a putative transmembrane domain is centered
XX      around this residue"
XX      150
XX      /note= "a putative transmembrane domain is centered
XX      around this residue"
XX      W09947555-A1.
XX      23-SEP-1999.
XX      18-MAR-1999; 99WO-US05939.
XX      20-MAR-1998; 98US-0078803.
XX      17-MAR-1999; 99US-0078803.
XX      (GENY ) GENETICS INST INC.
XX      PI      Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX      PI      Merberg D, Treacy M, Agostino MT, Steindinger RJ;
XX      DR      WPI; 1999-562059/47.
XX      DR      N-PSDB; AA219894.
XX      New polynucleotides derived from murine foetal cell cDNA libraries,
XX      potentially used as, e.g. vaccines
XX      Claim 13(a); Page 94; 107pp; English.
XX      This is the predicted amino acid sequence of a novel human secreted
XX      protein, g21_1, as deduced from an isolated adult blood cDNA

```

```

CC      clone (see AA219894). The invention provides new human secreted
CC      proteins (see AAY31828-38) and polynucleotides (see AA219893-901)
CC      isolated from foetal cell, adult blood, adult brain and foetal
CC      kidney cDNA libraries. They are predicted to have biological
CC      activities which would make them suitable for treating, preventing
CC      or ameliorating medical conditions in humans and animals, although
CC      no supporting data are given. Suggested activities include
CC      nutritional, cytokine, tissue growth, cell proliferation and
CC      differentiation, immunostimulant (e.g. as vaccine),
CC      immunosuppressive, haematopoiesis regulating, activin or inhibin,
CC      chemotactic or chemokinetic, haemostatic or thrombolytic,
CC      receptor/ligand activity, antiinflammatory, catheterin or tumour
CC      invasion suppressor, and tumour inhibition activities.
XX      SQ      Sequence 236 AA;
XX
XX      Query Match      99.4%; Score 1203; DB 20; Length 236;
XX      Best Local Similarity 99.6%; Pred. No. 9.4e-93;
XX      Matches 235; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 MLSKGLKRRREERKEPLAVDSWMLDPGHAAPVAVASSSLFDLSVLKHSIQDS 60
DB      1 MLSKGLKRRREERKEPLAVDSWMLDPGHAAPVAVASSSLFDLSVLKHSIQDS 60
QY      61 EEDLRHLVLYVNTLRRIQASMAPPAALPPVSPPAAPSVADNLLASSDAALSASMASLIE 120
DB      61 EEDLRHLVLYVNTLRRIQASMAPPAALPPVSPPAAPSVADNLLASSDAALSASMASLIE 120
QY      121 DLSHIEGLSQAPQPLADEGPPGRSISGAAPSLGALDLGPATGCLLDGLEGLEFEDIDTS 180
DB      121 DLSHIEGLSQAPQPLADEGPPGRSISGAAPSLGALDLGPATGCLLDGLEGLEFEDIDTS 180
QY      181 MYDNEMLAPASBGLKPGPEDGPKKEAPLDEALDEILDYIMDVLYGTQALERPPGGR 236
DB      181 MYDNEMLAPASBGLKPGPEDGPKKEAPLDEALDEILDYIMDVLYGTQALERPPGGR 236

RESULT 6
AA302619
ID      AAY02619 standard; Protein; 236 AA.
XX
AC      AAY02619;
XX
DT      21-JUL-1999 (first entry)
XX
XX      Amino acid sequence of a p16 binding protein.
DE
XX      p16BP1; cell cycle regulator protein p16; binding protein;
XX      cell cycle regulation; cyclin D dependent kinase; Cdk4;
XX      cell cycle progression; 19q13.1-13.2; cancer; gene therapy;
XX      cell cycle disorder.
XX      Homo sapiens.
XX      W09923218-A1.
XX      14-MAY-1999.
XX      04-NOV-1998; 98WO-JP04972.
XX      05-NOV-1997; 97JP-0302564.
XX      (SOME ) SUMITOMO ELECTRIC IND CO.
XX      PI      Hara E, Nakamura T;
XX      DR      WPI; 1999-313340/26.
XX      DR      N-PSDB; AAX36177.
XX      New p16-binding protein useful in gene therapy of cancer and other
XX      cell cycle disorders
XX      Claim 1; Page 34-35; 40pp; Japanese.

```

XX The present sequence represents a protein, designated p16p1, which binds
 CC to the cell cycle regulator protein p16. p16p1 plays a role in cell
 CC cycle regulation by binding to p16 which is a protein binding to a
 CC cyclin D dependent kinase (CDK4) involved in the regulation of cell
 CC cycle progression. The gene encoding p16p1 has been localized to
 CC 19q13.1-13.2. The binding protein, polynucleotides encoding it, and
 CC antibodies recognizing it are useful in the investigation of cell cycle
 CC progression and regulation, particularly in connection with cancer
 CC formation, for gene therapy of cancer and other cell cycle disorders,
 CC and for diagnosis of such disorders. The sense or antisense
 CC polynucleotides are useful as probes to screen a cDNA library.

SQ Sequence 236 AA:
 Query Match 99.0%; Score 1198; DB 20; Length 236;
 Best Local Similarity 99.2%; Pred. No. 2.5e-92;
 Matches 234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLSKGLKRRREEEKEPLAVDSWMLDPGHAAVAQAAPVAVASSSLFDLSVLKLIHSLQGS 60
 DB 1 MLSKGLKRRREEEKEPLAVDSWMLDPGHAAVAQAAPVAVASSSLFDLSVLKLIHSLQGS 60
 QY 61 EEDLRHLVLYVNTLRRIQASMAPAALPPVPSPPAPSVADNULASSDAALSASMASLLE 120
 DB 61 EEDLRHLVLYVNTLRRIQASMAPAALPPVPSPPAPSVADNULASSDAALSASMASLLE 120
 QY 121 DLSHIEGLSQAPQPLADEGPPRSIGGAPSLGALDILGPATGCLLDLDEGLFEDIDTS 180
 DB 121 DLSHIEGLSQAPQPLADEGPPRSIGGAPSLGALDILGPATGCLLDLDEGLFEDIDTS 180
 QY 181 MTDNELMAPASGLKPGPEDGPGKEBAPLDEAELDYLMVDLVGTQALERRPPGGR 236
 DB 181 MTDNELMAPASGLKPGPEDGPGKEBAPLDEAELDYLMVDLVGTQALERRPPGGR 236

RESULT 7
 AA36004
 ID AAY36004 standard; Protein; 236 AA.
 AC AAY36004;
 XX
 XX 13-SEP-1999 (first entry)
 DE Extended human secreted protein sequence, SBO ID NO. 389.
 XX
 KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
 KW cellular differentiation; immune system regulator; anti-inflammatory;
 KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
 KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
 KW genetic disease.
 XX
 OS Homo sapiens.
 XX
 PN WO9931236-A2.
 PD 24-JUN-1999.
 XX
 PF 17-DEC-1998; 98WO-1B02122.
 XX
 PR 10-AUG-1998; 98US-0096116.
 PR 17-DEC-1997; 97US-0069957.
 PR 09-FEB-1998; 98US-0074121.
 PR 13-APR-1998; 98US-0081563.
 XX
 PA (GEST) GENSET.
 XX
 PI Bougueleret L, Duclet A, Dumas Milne Edwards J;
 XX
 DR WPI; 1999-385906/32.
 DR N-PSDB; AAX97688.
 XX
 PT New isolated human secreted proteins

XX Claim 9: Page 330-331; 516pp; English.
 PS
 XX This sequence is encoded by an extended human secreted protein coding
 CC sequence of the invention. The secreted proteins can be used in treating
 CC or controlling a variety of human conditions. The secreted proteins may
 CC act as cytokines or may affect cellular proliferation or differentiation
 CC or may act as immune system regulators, haematopoiesis regulators, tissue
 CC growth regulators, regulators of reproductive hormones or cell movement
 CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
 CC tumour inhibition activity. The DNAs can be used in forensic procedures
 CC to identify individuals or in diagnostic procedures to identify
 CC the genes corresponding to the extended cDNAs. They are also useful for
 CC constructing a high resolution map of the human chromosomes. They can
 CC also be used for gene therapy to control or treat genetic diseases.

SQ Sequence 236 AA:
 Query Match 96.2%; Score 1164; DB 20; Length 236;
 Best Local Similarity 95.8%; Pred. No. 1.7e-89;
 Matches 226; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 MLSKGLKRRREEEKEPLAVDSWMLDPGHAAVAQAAPVAVASSSLFDLSVLKLIHSLQGS 60
 DB 1 MLSKGLKRRREEEKEPLAVDSWMLDPGHAAVAQAAPVAVASSSLFDLSVLKLIHSLQGS 60
 QY 61 EEDLRHLVLYVNTLRRIQASMAPAALPPVPSPPAPSVADNULASSDAALSASMASLLE 120
 DB 61 EEDLRHLVLYVNTLRRIQASMAPAALPPVPSPPAPSVADNULASSDAALSASMASLLE 120
 QY 121 DLSHIEGLSQAPQPLADEGPPRSIGGAPSLGALDILGPATGCLLDLDEGLFEDIDTS 180
 DB 121 DLSHIEGLSQAPQPLADEGPPRSIGGAPSLGALDILGPATGCLLDLDEGLFEDIDTS 180
 QY 181 MTDNELMAPASGLKPGPEDGPGKEBAPLDEAELDYLMVDLVGTQALERRPPGGR 236
 DB 181 MTDNELMAPASGLKPGPEDGPGKEBAPLDEAELDYLMVDLVGTQALERRPPGGR 236

RESULT 8
 ABG14355
 ID ABG14355 standard; Protein; 116 AA.
 AC ABG14355;
 XX
 XX 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #14346.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS78542.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID NO 44714; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIRO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 116 AA;

Query Match 27.7%; Score 335.5; DB 22; Length 116;
Best Local Similarity 69.5%; Pred. No. 1.6e-20;
Matches 73; Conservative 5; Mismatches 14; Indels 13; Gaps 4;

OY 139 GPGRSIGGAAPSLGALDLGPAFGCLLDGLEGLEFEDIDTSMYDNE----LWAPASEG 193
DB 3 GPOASSIGGAAPSLGALDLGPAFGCLLDGLEGLEFEDIDTSMYDNEGLTSLWRASMQA 62

OY 194 LKPGEDGPG-KKEAPFELDEHEDLYMDVLYGTQAL---EPPPGP 234
DB 63 LK---DGPQGRKRLKSLDEAEELDYLMDEVLTGATGSDRGPPR 103

RESULT 9
AA12453
ID AA12453 standard; Protein: 58 AA.

AC AA12453;

DT 17-JUN-1999 (first entry)

DE Human 5' EST secreted protein SEQ ID NO:484.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX forensic; gene therapy; chromosome mapping; signal peptide;
XX upstream regulatory sequence; cytokine activity; cell proliferation;
XX differentiation; haematopoiesis regulation; tissue growth regulation;
XX reproductive hormone regulation; chemokine; chemokine; haemostatic;
XX thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.

PN WO906548-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; 98WO-IB01222.

PR 01-AUG-1997; 97US-0905135.

PA (BEST) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI: 1999-153778/13.

DR N-PSDB: AAX41286.

PT New nucleic acids encoding human secreted proteins - obtained from
PT CDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue

PS Claim 27; Page 782-783; 824pp; English.

XX AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
XX human secreted proteins, and encode the proteins given in AAX1261 to
XX AAX12514, respectively. The proteins given represent the signal peptide
XX and an N-terminal fragment of a secreted protein. The nucleic acid
XX sequences can be used for producing secreted human gene products. They
XX can also be used to develop products for diagnosis and therapy. The
XX proteins obtained may have cytokine activity, cell
XX proliferation/differentiation activity, haematopoiesis regulating
XX activity, tissue growth regulating activity, reproductive hormone
XX regulating activity, chemokine/chemokine activity, haemostatic and
XX thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, tumour inhibition activity or other activities. The products
XX can be used in forensic, gene therapy and chromosome mapping procedures.
XX The sequences can also be used for obtaining corresponding promoter
XX sequences. The nucleic acids encoding the signal peptide can be used for
XX detecting extracellular secretion of a polypeptide or the insertion of a
XX polypeptide into a membrane, or importing a polypeptide into a cell.

SQ Sequence 58 AA;

Query Match 23.9%; Score 289; DB 20; Length 58;
Best Local Similarity 100.0%; Pred. NO. 5.4e-17;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MLSKGLKRRKEEKEPELAVDSWMLDPGHAAPPAVASSSLFSLVYLKTHS 56
DB 1 MLSKGLKRRKEEKEPELAVDSWMLDPGHAAPPAVASSSLFSLVYLKTHS 56

RESULT 10
ID AAM93922 standard; Protein: 196 AA.

AC AAM93922;

DT 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 4085.

XX Human; full length cDNA; cDNA synthesis; oligo-capping.

OS Homo sapiens.

PN EP1130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

PA (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR WPI: 2001-524255/58.

DR N-PSDB: AAK94884.

PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

PS Claim 8; SEQ ID NO 4085; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' and 3' ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 196 AA;

Query Match 17.6%; Score 213; DB 22; Length 196;
Best Local Similarity 32.6%; Pred. No. 5.4e-10;
Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

OY 5 GLKRRKEEKEEPLAVDSM-WLDPGAAVAQAAPPAVASSSLFDLSVLKHSIQSEPD 63
DB 4 GLKRRHSDLEEBE-ERMWSPAGLOSTYQA-----LIRISLDKVSRLGRAPS 51
OY 64 LRHLVLVNTLRRIQAS--MAPAALPPVSPPAFSAVDNLLASSDAALSASMSLTD 121
DB 52 LRHHVLIHNTLQQLQALRLAPALPAPLPPL-----FLGEEDFSLSATIGSILRE 101
OY 122 L-SHIEGLSQAPQ-----LADGPPGRSTIGGAAPSLGALDLGPATGCLLDGLEGLE 175
DB 102 LDTSMGTEPPQNPVTPLGLQNEVPP-----QDPVFELEL--SSRYLGSGLDDEFL 152
OY 176 DIDTSMYDNELMAPASEGLKPGPED---GPKKEAPELDEAELDYLMADVGT 225
DB 153 DIDTSAVEKE---PARAPEP-PHNLFCAFGSWE-----WNELDHIMEITILGS 196

RESULT 11
ID AAB35402 standard; Protein; 196 AA.
XX
AC AAB35402;
XX
DT 23-MAY-2001 (first entry)
XX
DE Replication protein A binding transcriptional activator 1 RBT1.
XX
KM RBT1; replication protein A binding transcriptional activator 1;
KM RPA32; gene therapy; apoptosis; cancer; leukaemia.
XX
OS Unidentified.
XX
PN WO200114546-A2.
XX
PD 01-MAR-2001.
XX
PF 17-AUG-2000; 2000WO-CA00948.
XX
PR 19-AUG-1999; 99US-0149472.
XX
PA (TRAN-) CENT TRANSLATIONAL RES IN CANCER.
XX
PI Alaoui-Jamali MA, Cho JM;
XX
DR WPI: 2001-218447/22.
XX
PT N-PSDB; AAF28052.
XX
PT Novel replication protein A binding transcriptional activator 1 gene,
XX
PT useful for treating neoplastic disorders such as cancer and in gene
XX
PS therapy -
PS Disclosure: Fig 1; 16pp; English.

XX
CC The present invention provides the protein and coding sequences of the
CC replication protein A binding transcriptional activator 1 (RBT1). The
CC protein is capable of inducing apoptosis. The sequences are useful in the
CC gene therapy and other methods of treatment of cancer, including
CC leukaemias. The present sequence is the RBT1 protein.
XX
SQ Sequence 196 AA;

Query Match 17.6%; Score 213; DB 22; Length 196;
Best Local Similarity 32.6%; Pred. No. 5.4e-10;
Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

OY 5 GLKRRKEEKEEPLAVDSM-WLDPGAAVAQAAPPAVASSSLFDLSVLKHSIQSEPD 63
DB 4 GLKRRHSDLEEBE-ERMWSPAGLOSTYQA-----LIRISLDKVSRLGRAPS 51
OY 64 LRHLVLVNTLRRIQAS--MAPAALPPVSPPAFSAVDNLLASSDAALSASMSLTD 121
DB 52 LRHHVLIHNTLQQLQALRLAPALPAPLPPL-----FLGEEDFSLSATIGSILRE 101
OY 122 L-SHIEGLSQAPQ-----LADGPPGRSTIGGAAPSLGALDLGPATGCLLDGLEGLE 175
DB 102 LDTSMGTEPPQNPVTPLGLQNEVPP-----QDPVFELEL--SSRYLGSGLDDEFL 152
OY 176 DIDTSMYDNELMAPASEGLKPGPED---GPKKEAPELDEAELDYLMADVGT 225
DB 153 DIDTSAVEKE---PARAPEP-PHNLFCAFGSWE-----WNELDHIMEITILGS 196

RESULT 12
ID AAB65076 standard; Protein; 314 AA.
XX
AC AAB65076;
XX
DT 12-NOV-2002 (first entry)
XX
DE Hypoxia-induced protein #16.
XX
KM Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;
KM antiinflammatory; vulnerrary; gynecological; ophthalmological; vaccine;
KM hypoxia; tumourigenesis; angiogenesis; apoptosis; cancer;
KM ischaemic condition; reperfusion injury; retinopathy; neonatal stress;
KM precleamsia; atherosclerosis; inflammatory condition; wound healing;
KM inflammation; erythropoiesis; hair loss; human.
XX
OS Homo sapiens.
XX
PN WO200246465-A2.
XX
PD 13-JUN-2002.
XX
PF 10-DEC-2001; 2001WO-GB05458.
XX
PR 08-DEC-2000; 2000GB-0030076.
XX
PR 08-FEB-2001; 2001GB-0003156.
XX
PR 25-OCT-2001; 2001GB-0025666.
XX
PA (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
PI White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;
XX
PI Rayner WN;
XX
DR WPI: 2002-627238/67.
XX
PT Identifying a gene involved in disease for treating hypoxia-regulated
XX
PT conditions, comprises comparing the transcriptome/proteome of two cell
XX
PT types under different conditions and identifying a differentially
XX
PS regulated gene -
PS Claim 13; Page 280; 538pp; English.

[illegible]

| | | |
|---|--|--|
| Oy | | 115 M-----AALIEDLDSHIG-----LSAPPLADEG----PEPRISGGAAPISG- 153 |
| Dd | | 119 WGDEGAHPASGLAD -GHIOGPVSDLCPTVISAAPRHLDSAMEMDGRPNENKGSFHKSLDQ 177 |
| Oy | | 154 ---ALDLLGPAATCCLLDLDGLEGFEDIDTSMYD-NELMAPASEGLKPGEDEG----- 201 |
| Dd | | 178 IFETLETKNDS--C-----MEELFSVDSPYYDLDPVLTMGMGARPPCGBLGPLPAT 230 |
| Oy | | 202 PGKEAPELDELDAFLDYLVGT 225 |
| Dd | | 231 PGPSSCKSDLGELDHYVELIVET 254 |
| RESULT 15 AAB58258 ID AAB58258 standard; Protein; 254 AA. XX XX AAB58258; XX DT 14-MAR-2001 (first entry) DE Lung cancer associated polypeptide sequence SEQ ID 596. XX XX Human; lung cancer associated protein; neuroprotective; cytosolic; KW cardiocactive; immunomodulatory; muscular active; vulnerary; KW gastrointestinal; nephroproctic; antinfetive; gynecological; KM antibacterial; diagnos; neural disorder; immune disorder; reproductive; KW proliferative disorder; wound healing; infectious disease. XX XX Homo sapiens. OS PM WO20005180-A2. PD 21-SEP-2000. XX XX 08-MAR-2000; 2000WO-US05918. PF XX 12-MAR-1999; 99US-0124270. PR XX (HUMA-) HUMAN GENOME SCI INC. PA (ROSE/) ROSEN C A. XX XX Ruben SM; PI XX WPI: 2000-567514/55. DR N-PSDB; AAF18134. XX XX Lung cancer associated gene sequences, referred to as lung cancer PT antigens, useful for treatment, prevention, and diagnosis of disorders PT such as lung cancer - XX XX Claim 11: Page 1089-1090; 1425pp; English. | | |
| PS | | Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer |
| XX | | associated proteins represented in AAB58106 - AAB58548. Lung cancer |
| CC | | associated proteins and polynucleotide sequences, their agonists, and |
| CC | | antagonists may have neuroprotective; cytosolic; cardioactive; |
| CC | | immunomodulatory; muscular active general; vulnerary; gastrointestinal |
| CC | | general; nephroproctic; antinfetive; gynecological; or antibacterial |
| CC | | activity. The invention also includes antibodies specific for the |
| CC | | protein or polynucleotide sequences. The lung cancer associated |
| CC | | polynucleotide sequences may be used for detection of lung cancer, |
| CC | | chromosome identification, as chromosome markers, and for numerous other |
| CC | | diagnostic or research purposes. The proteins may be used to treat |
| CC | | disorders such as neural, immune, muscular, reproductive, |
| CC | | gastrointestinal, pulmonary, cardiovascular, renal, and proliferative |
| CC | | disorders. The proteins may also be used in the treatment of wounds and |
| CC | | infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and |
| CC | | peptide AAB58549 are used in the course of the invention for the |
| CC | | identification and characterisation of the polynucleotide and protein |
| XX | | sequences. |
| XX | | |

SQ Sequence 254 AA:

Query Match 13.9%; Score 168; DB 21; Length 254;
Best Local Similarity 27.7%; Pred. No. 4.3e-06;
Matches 73; Conservative 39; Mismatches 90; Indels 62; Gaps 13;

```
QY 1 MISKGLKRR--REEEEKEPLA----VDSWMLDPGHAAVAQAPPAVASSSLFDLSVKLH 54
DB 14 MFARGLKRRKCVGHEDEVEGALAGLKTVSSY-----SLQRSLIDMSLVKLDQ 59
QY 55 HSLQSEPDRLHLVVTNLRRIQASMAPAALPPVSPAPASVADNLLASDAALAS 114
DB 60 LCHMLVEPNLCRSVLIANTVROIQEEMTODGTWRTV-APQAARAPXDRIVSTETLCRAA 118
QY 115 M-----ASLLEDLSHIEG-----LSQAQOPLADEG---PGRSIGGAAPSLG- 153
DB 119 WGOEGAHAPAPGIGD-GHTQGPVSDLCPYTSAQAPRHLOSASAWEMDGPRENREGSFHKSLDQ 177
QY 154 ---AIDLGPATGCLLDGLEGLEDIDTSMYD-NELNAPASEGLKPGPEDG----- 201
DB 178 IFETLETENPS--C-----MEELFSDVDSPYYDLDITVLTGMWGGARPGPCBGLGLAPAT 230
QY 202 PQKEAPEIDEAEIDYLDMDVLVGT 225
DB 231 RGPSSCKSDGLGELDHVVEILVET 254
```

Search completed: August 19, 2003, 09:36:25
Job time : 87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:31:59 ; Search time 11 Seconds

(without alignments)
1008.937 Million cell updates/sec

Title: US-09-701-675A-3

Perfect score: 1210
Sequence: 1 MMSGKLRKREEREEKEPLA.....YIMDVLTGQALERPPGPR 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 208.5 | 17.2 | 314 | Y127_HUMAN | Q14140 homo sapien |
| 2 | 120 | 9.9 | 3164 | 1 TEGU_HSV11 | P10720 herpes simp |
| 3 | 118.5 | 9.8 | 2161 | 1 SHK1_HUMAN | Q9Y566 homo sapien |
| 4 | 112.5 | 9.3 | 652 | 1 DRBR_CHICK | P18302 gallus gall |
| 5 | 112 | 9.3 | 863 | 1 M113_HUMAN | Q8N3F8 homo sapien |
| 6 | 108.5 | 9.0 | 488 | 1 CTA0_HUMAN | Q96m4 homo sapien |
| 7 | 107.5 | 8.9 | 440 | 1 FXGA_CHICK | Q98337 gallus gall |
| 8 | 107.5 | 8.9 | 708 | 1 VP40_HCMVA | P16753 human cytom |
| 9 | 106.5 | 8.8 | 663 | 1 DUS8_MOUSE | O09112 mus musculu |
| 10 | 106.5 | 8.8 | 909 | 1 CNG4_HUMAN | Q14028 homo sapien |
| 11 | 105 | 8.7 | 447 | 1 ODP2_RHIME | Q919n3 rhizobium m |
| 12 | 103.5 | 8.6 | 1567 | 1 FNN2_MOUSE | Q91104 mus musculu |
| 13 | 102.5 | 8.5 | 901 | 1 A180_MOUSE | O61548 mus musculu |
| 14 | 102.5 | 8.5 | 908 | 1 SRC4_RABIT | P13666 oryctolagus |
| 15 | 102 | 8.4 | 1516 | 1 CA1H_HUMAN | P39060 homo sapien |
| 16 | 101.5 | 8.4 | 676 | 1 ICP0_HSVBK | P29836 bovine hepp |
| 17 | 101.5 | 8.4 | 2774 | 1 MAPA_RAT | P34926 rattus norv |
| 18 | 101 | 8.3 | 433 | 1 AFLR_EMENT | O15027 homo sapien |
| 19 | 101 | 8.3 | 1433 | 1 Y310_HUMAN | P29128 bovine hepp |
| 20 | 100.5 | 8.3 | 676 | 1 ICP0_HSVBJ | Q12778 bovine hepp |
| 21 | 100 | 8.3 | 655 | 1 FXO1_HUMAN | Q91wv8 mus musculu |
| 22 | 100 | 8.3 | 806 | 1 MK07_MOUSE | O91wv8 mus musculu |
| 23 | 100 | 8.3 | 1175 | 1 HCN4_RABIT | O91wv8 mus musculu |
| 24 | 100 | 8.3 | 1509 | 1 GSR1_HUMAN | O91wv8 mus musculu |
| 25 | 99.5 | 8.2 | 359 | 1 CEB4_MOUSE | P53566 mus musculu |
| 26 | 99.5 | 8.2 | 1106 | 1 IP0D_BOVIN | P28339 bos taurus |
| 27 | 99 | 8.2 | 373 | 1 HMT1_DROST | Q24C48 drosophila |
| 28 | 99 | 8.2 | 498 | 1 VGLX_PRIVI | P07562 pseudophila |
| 29 | 98.5 | 8.1 | 281 | 1 PMXA_RAT | O62782 rattus norv |
| 30 | 98.5 | 8.1 | 2167 | 1 SHK1_RAT | O9w48 rattus norv |
| 31 | 98 | 8.1 | 1199 | 1 J1P2_HUMAN | O13387 homo sapien |
| 32 | 98 | 8.1 | 1324 | 1 P121_RAT | P52591 rattus norv |
| 33 | 98 | 8.1 | 3010 | 1 POLG_HCVTV | P29846 h genome po |

| | | | | | |
|----|------|-----|------|--------------|--------------------|
| 34 | 97.5 | 8.1 | 245 | 1 VE4_HPV05 | P06924 human papil |
| 35 | 97.5 | 8.1 | 1089 | 1 Y553_HUMAN | Q9ukf3 homo sapien |
| 36 | 97 | 8.0 | 519 | 1 IRX4_HUMAN | P78413 homo sapien |
| 37 | 97 | 8.0 | 2333 | 1 PGCA_CANFA | Q28343 canis fam1 |
| 38 | 96.5 | 8.0 | 1324 | 1 IRS2_HUMAN | Q9Y4h2 canis fam1 |
| 39 | 96.5 | 8.0 | 1527 | 1 CA1H_MOUSE | P39061 mus musculu |
| 40 | 96.5 | 8.0 | 2067 | 1 NC06_MOUSE | Q9j119 mus musculu |
| 41 | 96 | 7.9 | 340 | 1 GBX2_CHICK | O42230 gallus gall |
| 42 | 96 | 7.9 | 439 | 1 VSX1_CHICK | O91a12 gallus gall |
| 43 | 96 | 7.9 | 350 | 1 XP2_XENLA | P17437 xenopus lae |
| 44 | 96 | 7.9 | 564 | 1 ARX_MOUSE | O35085 mus musculu |
| 45 | 96 | 7.9 | 1298 | 1 ICP4_HSV11 | P08392 herpes simp |

ALIGNMENTS

| | | | | | |
|-----------------------|---|------------|-----------|------|---------|
| RESULT 1 | ID | Y127_HUMAN | STANDARD; | PRT; | 314 AA. |
| AC | Q14140; | | | | |
| DT | 15-JUL-1998 (Rel. 36, Created) | | | | |
| DT | 15-JUL-1998 (Rel. 36, Last sequence update) | | | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | | | |
| DE | Hypothetical protein KIAA0127. | | | | |
| GN | KIAA0127. | | | | |
| OS | Homo sapiens (Human). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| OX | NCBI_TaxID=9606; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | TISSUE=bone marrow; | | | | |
| RX | MEDLINE=96127530; PubMed=8590280; | | | | |
| RA | Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.; | | | | |
| RT | "Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by analysis of cDNA clones from human cell line KG-1." | | | | |
| RL | DNA Res. 2:167-174(1995). | | | | |
| CC | ----- | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib.ch). | | | | |
| CC | or send an email to license@sib.ch . | | | | |
| DR | EMBL; D50917; BAA09476.1; .. | | | | |
| KW | Hypothetical protein. | | | | |
| SK | SEQUENCE 314 AA; 33896 MW; E43107FC565AAC31 CRC64; | | | | |
| Query Match | 17.2%; Score 208.5; DB 1; Length 314; | | | | |
| Best Local Similarity | 27.2%; Pred. No. 3.7e-07; | | | | |
| Matches | 88; Conservative 37; Mismatches 91; Indels 107; Gaps 11; | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| DB | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA-----SSLPDLSVLKHS 56 | | | | |
| OY | 1 MMSGKLRKREEREEKEPLAVDSWMLDPGAAVAAQAPVVA | | | | |

```

DB      232 TTSTGFLDRLDLDLFDADIDISMDPCTSSSGTASMAVSAADLKLTAAPSQPV 291
OY      209 -----ELDEALDYIMDVLC 225
DB      292 TPSOPFKMDLFEIDHIMEVLVS 314

RESULT 2
TEGU_HSV11 STANDARD: PRT: 3164 AA.
ID      TEGU_HSV11
AC      P10220:
DT      01-MAR-1989 (Rel. 10, Created)
DT      01-MAR-1989 (Rel. 10, Last sequence update)
DT      01-APR-1993 (Rel. 25, Last annotation update)
DE      Large tegument protein (Virion protein UL36).
GN      UL36.
OS      Herpes simplex virus (type 1 / strain 17).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Alphaherpesvirinae; Simplexvirus.
OX      NCBI_TaxID=10299;
RN      [1]
RX      MEDLINE=88274327; PubMed=2839594;
RA      McGeoch D.J., Dairymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA      McNaughton D., Perry L.J., Scott J.E., Taylor P.;
RT      "The complete DNA sequence of the long unique region in the genome of
RT      herpes simplex virus type 1."
RT      J. Gen. Virol. 69:1531-1574(1988).
CC      -1- FUNCTION: TEGUMENT PROTEIN.
CC      -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC      HSV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X14112; CA932311.1;
DR      PIR: I30085; MMBE6.
DR      InterPro: IPR006928; Herpes_teg_N.
DR      InterPro: IPR005210; Herpes_UL36.
DR      Pfam: PF04843; Herpes_teg_N; 1.
DR      Pfam: PF03586; Herpes_UL36; 1.
DR      Repeat.
FT      DOMAIN 2911 2980 35 X 2 AA TANDEM REPEATS OF P-Q.
SQ      SEQUENCE 3164 AA; 335857 MW; CC5D31FF4F9FE3F4 CRC64;

Query Match 9.9%; Score 120; DB 1; Length 3164;
Best Local Similarity 26.7%; Pred. No. 3.2;
Matches 65; Conservative 23; Mismatches 107; Indels 48; Gaps 10;

OY      21 VDSMWLDPBGHAAVQAAPVAVSSSLFDLSVCLKHSHQDSRPDLRHLYVNTLRIOAS 80
DB      2396 IETVAVRLHADLPNENACLAQDLPRLSALIAERPLAKGP-----DIS 2445
OY      81 MAPAAALPVVSPPAASV-----ADNLASS-----DAALSAMASILE 120
DB      2446 MTPVAVLMENDDPPGPPVRRVVGSEATELRFVAGGEVLAASATDEPFLARAILGKPF 2505
OY      121 DLSHIEGLSQAPPLADEGPPGSGSIGAAPSIGALD-----LGPAN--GCLLDGLEG 172
DB      2506 DASLISELFGHGVYVYGRAPDDQSPVNPTRPGVDVYAGSGSLAPTLFTDAVPG 2565
OY      173 LFEIDISMYNEMMAVASEGLAKRPEDGPKERAPELDEALDYIMDVLCGT-QALERP 231
DB      2566 --EVPVPMW--AWIHGLEELIASDGGP-----APLIAADPLSPTAQSVPTSQCAPRP 2616
OY      232 GGP 234
DB      III

```

```

DB      2617 GGP 2619

RESULT 3
SHK1_HUMAN STANDARD: PRT: 2161 AA.
ID      SHK1_HUMAN
AC      Q9Y566; Q9NYW9;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      SH3 and multiple ankyrin repeat domain protein 1 (Shank1)
DE      (Somatostatin receptor interacting protein) (SSTR interacting protein)
DE      (SSTRIP).
GN      SHANK1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RX      MEDLINE=20267867; PubMed=10806096;
RA      Sheng M., Kim E.;
RT      "The Shank family of scaffold proteins."
RT      J. Cell Sci. 113:1851-1856(2000).
CC      -1- FUNCTION: Seems to be an adapter protein in the postsynaptic
CC      density (PSD) of excitatory synapses that interconnects receptors
CC      of the postsynaptic membrane including NMDA-type and metabotropic
CC      glutamate receptors via complexes with GSK-3β and Homer,
CC      respectively, and the actin-based cytoskeleton. May play a role in
CC      the structural and functional organization of the dendritic spine
CC      and synaptic junction.
CC      -1- SUBUNIT: May homomultimerize via its SAM domain (By similarity).
CC      Interacts with SSTR2 C-terminus via the PZ domain. Interacts with
CC      SPTAN1, Homer-1 and DGAP1/GKAP isoforms 1 and 2 (By similarity).
CC      Is part of a complex with DLG4/PSD-95 and DLGAP1/GKAP (By
CC      similarity).
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
CC      neuronal cells (By similarity).
CC      -1- ALTERNATIVE PRODUCTS:
CC      Event-Alternative splicing; Named isoforms=3:
CC      Name=1: Synonyms=A;
CC      IsoId=Q9Y566-1; Sequence=Displayed;
CC      Name=2: Synonyms=B;
CC      IsoId=Q9Y566-2; Sequence=VSP_006069, VSP_006070;
CC      Name=3;
CC      IsoId=Q9Y566-3; Sequence=VSP_006071;
CC      -1- TISSUE SPECIFICITY: Expressed in brain particularly in the
CC      amygdala, hippocampus, substantia nigra and thalamus. Isoform 2
CC      seems to be expressed ubiquitously.
CC      -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.
CC      -1- SIMILARITY: Contains 6 ANK repeats.
CC      -1- SIMILARITY: Contains 1 PZ/DHR domain.
CC      -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC      -1- SIMILARITY: Contains 1 SH3 domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: AF163302; AAD45121.1;
DR      EMBL: AF226728; AAF35887.1;

```

| | |
|-----------------------|--|
| DR | HSSP; P06241; 1SHF. |
| DR | Gnew; HGNC:15474; SHANK1. |
| DR | MIM: 604999; - |
| DR | InterPro; IPR002110; ANK. |
| DR | InterPro; IPR001478; PDZ. |
| DR | InterPro; IPR001660; SAM. |
| DR | InterPro; IPR001452; SH3. |
| DR | Pfam; PF00023; ank; 7. |
| DR | Pfam; PF00595; pdz; 1. |
| DR | Pfam; PF00536; sam; 1. |
| DR | Pfam; PF00018; sh3; 1. |
| DR | ProDom; PD000066; SH3; 1. |
| DR | SMART; SM00248; ANK; 6. |
| DR | SMART; SM00228; PDZ; 1. |
| DR | SMART; SM00454; SAM; 1. |
| DR | SMART; SM00376; SH3; 1. |
| DR | PROSITE; PS50086; ANK_REPEAT; 3. |
| DR | PROSITE; PS50297; ANK_REP_REGION; 1. |
| DR | PROSITE; PS50106; PDZ; 1. |
| DR | PROSITE; PS50105; SAM_DOMAIN; 1. |
| DR | PROSITE; PS50002; SH3; 1. |
| KW | ANK repeat; SH3 domain; Repeat; Alternative splicing. |
| FT | REPEAT 212 245 |
| FT | REPEAT 246 278 |
| FT | REPEAT 279 312 |
| FT | REPEAT 313 345 |
| FT | REPEAT 346 378 |
| FT | REPEAT 379 395 |
| FT | DOMAIN 554 613 |
| FT | DOMAIN 663 757 |
| FT | DOMAIN 2098 2161 |
| FT | DOMAIN 1002 1007 |
| FT | DOMAIN 1014 1019 |
| FT | DOMAIN 1189 1195 |
| FT | DOMAIN 1709 1717 |
| FT | DOMAIN 1844 1854 |
| FT | DOMAIN 1896 1902 |
| FT | DOMAIN 1970 1979 |
| FT | VARSPLIC 1 613 |
| FT | VARSPLIC 614 654 |
| FT | VARSPLIC 646 654 |
| FT | VARSPLIC 646 654 |
| FT | SEQUENCE 2161 AA; 225019 MW; 5FECC969CBE98701 CRC64; |
| Query Match | 9.88; Score 118.5; DB 1; Length 2161; |
| Best Local Similarity | 26.78; Pred. No. 2.7; Mismatches 106; Indels 55; Gaps 10; |
| Matches | 66; Conservative 20; |
| DQ | 10 REEEEREKLAVDSDMWLDPGHA-----VAOAPPAVASSSLFDLSVLKLNHSDQ 59 |
| DQ | 1535 RASENGPLLY-----LPRAPSVDVEDGEFLFEPRLPPLEFSNFKSPSPILTPGRPH 1590 |
| DQ | 60 SEEDLRHLVYLVTNRKRQAASMAPAALPPVPSP--PAAPSADNLAAASDAALSAMAS 117 |
| DQ | 118 LLEDLSHIEGLSAPOLADGPCGRSI--GGAAPSLGLDLISPGATGCLLDGLEG--- 173 |
| DQ | 1591 PLRP-----TPRATPLPPPPPAVAAPPTLDS-TASSLTYSDEEAT 1633 |
| DQ | 1534 L-----TQGSAAAPGDHPHPGPAPAAPAAAPQPC---DPPPG--TDSGLEVDNR 1681 |
| DQ | 174 -----FEDIDTSMYDNELMWPASEGLKPGPEDSGKEADELDEALDYLMVLYGTQA 227 |
| DQ | 1682 SSDHPLETISSASTLSLAEGGGAGGGGAGVASGPELLDTYAYILDGAFGSS 1741 |
| DQ | 228 LRRPGR 234 |
| DQ | 1742 TPGPPYP 1748 |

| RESULT 4 | DREB_CHICK | STANDARD: | PRT: | 652 AA. |
|----------|--|-----------|------|---------|
| ID | DREB_CHICK | | | |
| AC | P18302; Q91358; Q91359; | | | |
| DT | 01-NOV-1990 (Rel. 16, Created) | | | |
| DR | 01-NOV-1997 (Rel. 35, Last sequence update) | | | |
| DE | 28-FEB-2003 (Rel. 41, Last annotation update) | | | |
| GN | Drebrin (Developmentally regulated brain protein). | | | |
| OS | Drebrin (Developmentally regulated brain protein). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| CC | Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; | | | |
| CC | Gallus. | | | |
| CC | NCBI_TaxID=9031; | | | |
| CC | [1] | | | |
| CC | SEQUENCE FROM N.A. (ISOFORMS E1 AND E2). | | | |
| CC | TISSUE=Brain: | | | |
| CC | MEDLINE=89089137; PubMed=3208110; | | | |
| CC | Kojima N., Kato Y., Shirao T., Odata K.; | | | |
| CC | "Nucleotide sequences of two embryonic drebrins, developmentally | | | |
| CC | regulated brain proteins, and developmental change in their mRNAs.;" | | | |
| CC | Brain Res. 464:207-215(1988). | | | |
| CC | [2] | | | |
| CC | SEQUENCE FROM N.A. (ISOFORM A). | | | |
| CC | MEDLINE=93368392; PubMed=8361332; | | | |
| CC | Kojima N., Shirao T., Odata K.; | | | |
| CC | "Molecular cloning of a developmentally regulated brain protein, | | | |
| CC | chicken drebrin A and its expression by alternative splicing of the | | | |
| CC | drebrin gene.;" | | | |
| CC | Brain Res. Mol. Brain Res. 19:101-114(1993). | | | |
| CC | -1- FUNCTION: DREBRINS MIGHT PLAY SOME ROLE IN CELL MIGRATION, | | | |
| CC | EXTENSION OF NEURONAL PROCESSES AND PLASTICITY OF DENDRITES, | | | |
| CC | RESPECTIVELY. BINDS F-ACTIN. | | | |
| CC | -1- SUBCELLULAR LOCATION: Cytoplasmic. | | | |
| CC | -1- ALTERNATIVE PRODUCTS: | | | |
| CC | Name=A; | | | |
| CC | Isoid=P18302-1; Sequence=Displayed; | | | |
| CC | Name=E1; | | | |
| CC | Isoid=P18302-2; Sequence=VSP_004196; | | | |
| CC | Name=E2; | | | |
| CC | Isoid=P18302-3; Sequence=VSP_004197; | | | |
| CC | -1- TISSUE SPECIFICITY: BRAIN NEURONS. | | | |
| CC | -1- MISCELLANEOUS: DREBRINS ARE CLASSIFIED INTO TWO FORMS OF THE | | | |
| CC | EMBRYONIC TYPE (E1 AND E2) AND ONE FORM OF THE ADULT TYPE (A). THE | | | |
| CC | TIME COURSE OF THEIR APPEARANCE ARE DIFFERENT FROM EACH OTHER. | | | |
| CC | THEIR STRUCTURES ARE CLOSELY RELATED. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| CC | EMBL; M36961; AAA48750.1; -; | | | |
| CC | EMBL; S65296; AAB28012.1; -; | | | |
| CC | EMBL; S65279; AAB28012.1; JOINED. | | | |
| CC | EMBL; S65280; AAB28012.1; JOINED. | | | |
| CC | EMBL; S65281; AAB28012.1; JOINED. | | | |
| CC | EMBL; S65286; AAB28012.1; JOINED. | | | |
| CC | EMBL; S65289; AAB28012.1; JOINED. | | | |
| CC | EMBL; S65290; AAB28012.1; JOINED. | | | |
| CC | EMBL; S65291; AAB28012.1; JOINED. | | | |
| CC | EMBL; S65292; AAB28012.1; JOINED. | | | |
| CC | EMBL; S65294; AAB28012.1; JOINED. | | | |
| CC | EMBL; S65297; AAB28011.2; -; | | | |
| CC | EMBL; S65230; AAB28010.1; -; | | | |
| CC | InterPro: IP0002108; Actindn_cofin. | | | |
| CC | Pfam: PF00241; cofilin_ADF.1. | | | |
| CC | SMART: SM00102; ADF.1. | | | |
| CC | Actin-binding; Neurone; Alternative splicing. | | | |

[illegible]

CC CC
RA Odebi C.N., Pavlitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soederlund C., Spraggon L., Steward C.A., Sutton J.E., Swann R.M.,
RA Vaidin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilmshurst L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Mimosima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Domman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malat J., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky J., Robinson T.,
RA Schiet P., Walker C., Wamsley P., Wolfdamm P., Pepin K., Nelson J.,
RA Koff I., Bedell J.A., Hillier L., Marais E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA McEwen H.E., Johnson A., Wong A.C.C., Morrow J.P., Edelman L.,
RA Kim U.J., Shizuya H., Simon M.I., Dunham J.B., Peyrard M., Kedira D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkinson P., Bodenleitch A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tiliahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [3]
RP SEQUENCE OF 25-863 FROM N.A.
RC TISSUE-Melanoma:
RA Ansoy W., Winkner U., Mewes H.-W., Weill B., Wiemann S.;
RL submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 73-863 FROM N.A.
RC TISSUE-Brain:
RX MEDLINE=21156230; PubMed=11258795;
RA Hirosewa M., Nagase T., Murahashi Y., Kitano R., Ohara O.;
RT "Identification of novel transcribed sequences on human chromosome 22
RT by expressed sequence tag mapping.";
RL DNA Res. 8:1-9(2001).
RN [5]
RP SEQUENCE OF 514-863 FROM N.A.
RC TISSUE-Choriocarcinoma:
RX MEDLINE=22388257; PubMed=12477937;
RA Klausberg R.L., Fellngold E.A., Grouse L.H., Derge J.G.,
RA Krausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Butelw K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Maruska K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.P., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshilyukki S., Carninci P., Prange C.,
RA Rosa S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bobak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko I., Boulford G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield J.S.N., Krzyzanski M.I., Skalska U., Smallus D.E.,
RA Schencher A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SUBUNIT: Blinds to Rab13.
CC -1- SIMILARITY: Contains 1 calpain-in-homology (CH) domain.
CC -1- SIMILARITY: Contains 1 LIM zinc-binding domain.
CC -1- CAUTION: Ref.3 (CAD39036) sequence differs from that shown due to
CC frameshifts in positions 486 and 507.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AJ496196; CAD42713.1; -
 DR EMBL: AL022311; CAB62980.1; -
 DR EMBL: AL022311; CAB62979.1; -
 DR EMBL: AL022311; CAB62978.1; ALT_SEQ.
 DR EMBL: AL834373; CAD39036.1; ALT_FRAME.
 DR EMBL: AL833860; CAD38718.1; -
 DR EMBL: AB051455; BAB33338.1; -
 DR EMBL: BC010900; AAH01090.1; ALT_INT.
 DR HSSP: C01082; 1BKR.
 DR InterPro: IPR001715; Calpantin-like.
 DR InterPro: IPR001781; LIM.
 DR Pfam: PF00412; LIM.1.
 DR Pfam: PF00412; LIM.1.
 DR SMART: SM00094; LIM.1.
 DR SMART: SM00033; CH.1.
 DR PROSITE: PS00021; CH.1.
 DR PROSITE: PS00478; LIM_DOMAIN_1; 1.
 DR PROSITE: PS00023; LIM_DOMAIN_2; 1.
 DR LIM domain; Metal-binding; Zinc.
 FT DOMAIN 2 105 CH.
 FT DOMAIN 162 225 LIM.
 FT DOMAIN 253 667 PRO-RICH.
 FT DOMAIN 428 437 POLY-GU.
 FT CONFLICT 137 137 P -> S (IN REF. 4).
 FT CONFLICT 212 212 F -> Y (IN REF. 3; CAD39036).
 FT CONFLICT 519 519 A -> S (IN REF. 3).
 SQ SEQUENCE 863 AA; 93441 MW; 755E3B57C6037292 CRC64;

Query Match 9.3%; Score 112; DB 1; Length 863;
 Best Local Similarity 26.1%; Pred. No. 2.6;
 Matches 71; Conservative 30; Mismatches 103; Indels 68; Gaps 16;

QY 3 SNGLRKRREERKE-----PLAVDSMW-LDPGHAIVAQAQPAV 40
 DB 421 SKYPNPFEEEDKKEEAPAPSLATSPALGHPESTPKSLHWYGITTPSSKTKRRAP 480
 QY 41 ASSSFLDLSVTLKHS-LQOSEPD--LRHLVLYVTLRRTQASMAAAL---PPVP--- 91
 DB 481 RAPS---ASPLALHASRLSHSEPPSPATSPALSVESLSSESASQTAGAELLEPPAPVPS 537
 QY 92 SPPA--APSVADNLLASSDPAALASMASLLEDLSHIEGISO-----APQPLADEGP-PCR 143
 DB 538 SEPAYHAQCTPCNPVPSLTNSSLASSELVE--PVEQMPQASPGIAPRTGSSGPPQPK 595
 QY 144 SIGGAPSLGALDLGPATGCLLDLQLEDFEDIDTSMYDNELMAPASEGLKPGPEDPG 203
 DB 596 PCSGATPT--PLLVGDRSP--VPSGSSSPQGVKSSCKENF-----NRKPSAASPA 646
 QY 204 KEAPELDEAEIDYLVAVGQALERRPPG 235
 DB 647 TKKATK-----GSKPV-RPPAPG 663

RESULT 6
 CTAO_HUMAN STANDARD; PRT; 488 AA.
 AC Q96NM4; Q96IG9; Q9BONS;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein C20orf100.
 GN C20ORF100.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_TaxID=9606;

RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,
 RA Matsumoto K., Hirano S., Sano S., Nomura R., Yoshikawa Y.,
 RA Matsumura Y., Moriya S., Chiba E., Momiyama H., Onogawa S.,
 RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R.,
 RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,
 RA Matenabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Ota T.,
 RA Makemata A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K.,
 RA Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y.,
 RA Sekine M., Kikuchi H., Kanda K., Wagatsuna M., Muraoka K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahari K., Maunio Y., Nagai K., Isegai T.,
 RT "NEO human cDNA sequencing project."
 RL Submitted (Oct-2001) to the EMBL/Genbank/DBD databases.

RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Mathews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill M.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahnam D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G., Lawlor S.,
 RA Lehaesialho M.H., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.D., McConachie L.J., McIay K., Mcmurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prithalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Seha H.K., Showkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracer A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilmung L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RT Rogers J.;
 RL "The DNA sequence and comparative analysis of human chromosome 20."
 RN Nature 414:865-871(2001).

RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Muscle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguigliano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.T., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;

```

CC CC      Isoid-Q96NM4-1; Sequence-Displayed;
CC CC      Name=2;
CC CC      Isoid-Q96NM4-2; Sequence-VSP_002187;
CC CC      Note-No experimental confirmation available;
CC CC      -1- SIMILARITY: Contains 1 HMG box domain.
CC CC      -1- CAUTION: It is uncertain whether Met-1 or Met-52 is the initiator.
CC CC      -1- CAUTION: Ref.1 (CAC36288) sequence differs from that shown due to
CC CC      erroneous gene model prediction.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
DR DR      EMBL; AK051313; BAB70860.1; -
DR DR      EMBL; AL034419; CAD43476.1; -
DR DR      EMBL; AL034419; CAD43477.1; -
DR DR      EMBL; AL12187; CAC36288.1; ALT-SEQ.
DR DR      EMBL; BC007636; AAH07636.1; -
DR DR      Genew; HGNC:16095; C20orf100.
DR DR      InterPro: IPR000910; HMG_12_Box.
DR DR      Pfam: PF00505; HMG-box; 1.
DR DR      PROSITE; PS50118; HMG_BOX_2; 1.
DR DR      Nuclear protein; DNA-binding; Alternative splicing.
FT FT      DNA_BIND 255 323
FT FT      DOMAIN 245 250
FT FT      PRO-RICH.
FT FT      VARSPLIC 302 302
FT FT      Isoform 2).
FT FT      /FTid-VSP_002187.
FT FT      D -> N (in Ref. 1).
SQ SEQUENCE 488 AA; 51604 MW; 687FD144CF30731A CRC64;

Query Match
Best Local Similarity 26.0%; Score 108.5; DB 1; Length 488;
Matches 58; Conservative 28; Mismatches 90; Indels 47; Gaps 10;

QY 7 KRRKEEKEPEPLAVDSMWLDPGHAAYQAAPAVASSSLPDLVLT-----KLHNSLOOS 60
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 247 KKKKDPPEQKPVSAVALFFPDQALIKGNP---SATFGVSKIVASMSDLSDEQKOS 303
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 61 EPLRLHLYLVYNTLRIRIQASMAPPAALPPVPSPAASVA-----DNILASSDALASAM 115
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 304 SPD-----QGETKSTQAN-PPAKMLPPQPMYAMGSLFTLPSPDLQAFRSGASPASL 355
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 116 ASLLEDLSHIEGLSQAPPLADEGPPGSRISGAPSL-GALDLDGPATGCLLDGLEGLE 174
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 356 ARTLGSGSLRPLGLSASPP-----PPSPFL---SPTLHQQLSLPPHAGALLS----- 400
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 175 EDIDTSMYDNLMLAPASEGK-----PGPEDGKGKEAP 208
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 401 PAVSMSPAPQPPVLPPTPMALQVOLAMSPSPQDPFHISEFP 443
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 7
FXGA_CHICK STANDARD; PRT; 440 AA.
AC 098937;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Forkhead box protein G1A (forkhead-related protein FKHL2)
DE (Transcription factor BF-2) (Brain factor 2) (BF2) (CBF-2) (T-14-6).
GN FOXG1A OR FKHL2 OR HFBF2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_Taxid=9031;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn; TISSUE=Retina;
RX MEDLINE=96338226; PubMed=8757134;
RA Yuasa J., Hirano S., Yamagata M., Noda M.;
RT "Visual projection map specified by topographic expression of
RT transcription factors in the retina."
RL Nature 382:632-635(1996).
CC CC      -1- FUNCTION: MAY DETERMINE THE NASOTEMPORAL AXIS OF THE RETINA, AND
CC CC      CONSEQUENTLY SPECIFY THE TOPOGRAPHICAL PROJECTION OF THE RETINAL
CC CC      GANGLION-CELL AXONS TO THE TECTUM BY CONTROLLING EXPRESSION OF
CC CC      THEIR TARGET GENES.
CC CC      -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC CC      -1- TISSUE SPECIFICITY: Retina.
CC CC      -1- DEVELOPMENTAL STAGE: CAN BE DETECTED IN REGIONS INCLUDING
CC CC      PRIMITIVE RETINA AND NEUROEPITHELIUM BY EMBRYONIC DAY 2 (E2). AT
CC CC      E3, EXPRESSED IN THE TEMPORAL RETINA AND ASSOCIATED PIGMENT
CC CC      EPITHELIUM AS WELL AS IN PART OF THE DIENCEPHALON, AND AT E7 IS
CC CC      EXPRESSED IN RETINAL GANGLION CELLS. LEVELS BEGIN TO DECLINE FROM
CC CC      E4 AND ALMOST DISAPPEAR BY E10.
CC CC      -1- SIMILARITY: Contains 1 fork-head domain.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
DR DR      EMBL; U47276; AAB08467.1; -
DR DR      PIR; S71795; S71795.
DR DR      HSSP; Q63245; ZHPH.
DR DR      InterPro: IPR001766; TF_Fork_head.
DR DR      Pfam; PF00250; Fork_head; 1.
DR DR      PRINTS; PR00053; FORKHEAD.
DR DR      ProDom; PD000425; TF_Fork_head; 1.
DR DR      SMART; SMO0339; FH.1.
DR DR      PROSITE; PS00657; FORK_HEAD_1; 1.
DR DR      PROSITE; PS00658; FORK_HEAD_2; 1.
DR DR      PROSITE; PS50039; FORK_HEAD_3; 1.
DR DR      Transcription regulation; DNA-binding; Nuclear protein;
DR DR      developmental protein.
FT FT      DOMAIN 39 43
FT FT      POLY-ARG.
FT FT      DOMAIN 49 58
FT FT      POLY-GLU.
FT FT      DOMAIN 77 88
FT FT      POLY-GLY.
FT FT      DOMAIN 109 112
FT FT      POLY-ALA.
FT FT      DOMAIN 113 137
FT FT      POLY-GLY.
FT FT      DNA_BIND 143 237
FT FT      FORK-HEAD.
FT FT      DOMAIN 321 330
FT FT      POLY-PRO.
SQ SEQUENCE 440 AA; 44669 MW; 8226C2E1E103A48 CRC64;

Query Match
Best Local Similarity 32.3%; Score 107.5; DB 1; Length 440;
Matches 40; Conservative 12; Mismatches 45; Indels 27; Gaps 6;

QY 82 AAPAALPPVSPPAASVADNLASSDA-----LSASMSLLEDLSHIEGLSQAP 132
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 315 AAPAALPPPPPPPPPPRRAPLPPAELATPPGYPPLLPALAAASHAAKPPSGAAVAR 374
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 133 OPTADEGPPGSRISGAPSLGALDLDGPATG---CLLDG-LEGLEIDTSMYDNLML 167
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 375 SPFSIESITG---GGPFGIGA-----GPAFGAGSCASGAGATGSLSGSL----- 421
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 188 APAS 191
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 422 APAA 425

RESULT 8
VP40_HCMVA STANDARD; PRT; 708 AA.
ID VP40_HCMVA
AC P16753; 069030;
DT 01-AUG-1990 (Rel. 15, Created)

```


DT 01-ANG-1990 (Rel. 15, last sequence update)
 DT 15-SEP-2003 (Rel. 42, last annotation update)
 DE Capsid protein P40 [contains: Assemblin (Protease) (EC 3.4.21.97);
 DE Capsid assembly protein].
 GN UL80 OR APNG.
 OS Human cytomegalovirus (strain AD169).
 OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Cytomegalovirus.
 NCBI_Taxid=10360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90269039; PubMed=2161319;
 RA Chee M.S., Bankier A.T., Beck S., Bohnl R., Brown C.M., Cerny R.,
 RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
 RA Herdell E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.,
 RT "Analysis of the protein-coding content of the sequence of human
 RT cytomegalovirus strain AD169."
 RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF PROTEASE.
 RX MEDLINE=96399135; PubMed=8805706;
 RA Tong L., Qian C., Massariol M.-J., Bonneau P.R., Cordingley M.G.,
 RA Lagace L.,
 RT "A new serine-protease fold revealed by the crystal structure of
 RT human cytomegalovirus protease."
 RL Nature 383:272-275(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF PROTEASE.
 RX MEDLINE=96399136; PubMed=8805707;
 RA Olu X., Culp J.S., Dillella A.G., Hellmig B., Hoog S.S., Janson C.A.,
 RA Smith W.W., Abdel-Meguid S.A.,
 RT "Unique fold and active site in cytomegalovirus protease."
 RL Nature 383:275-279(1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.27 ANGSTROMS) OF PROTEASE.
 RX MEDLINE=96399137; PubMed=8805708;
 RA Shieh H.-S., Kurumbail R.G., Stevens A.M., Stegeman R.A.,
 RA Sturman E.J., Pak J.Y., Wiltner A.J., Palmer M.O., Wiegand R.C.,
 RA Holwerda B.C., Stallings W.C.,
 RT "Three-dimensional structure of human cytomegalovirus protease."
 RL Nature 383:279-282(1996).
 CC -1- FUNCTION: THE CAPSID ASSEMBLY PROTEIN IS A COMPONENT OF THE CAPSID
 CC CORE INVOLVED IN PROCESSING AND PACKAGING OF PROGENY DNA.
 CC ASSEMBLIN IS A PROTEASE WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF
 CC AND THE CAPSID ASSEMBLY PROTEIN AT THE C-TERMINUS.
 CC -1- CATALYTIC ACTIVITY: Cleaves -Ala-1-Ser- and -Ala-1-Ala- bonds in
 CC the scaffold protein.
 CC -1- PTM: CAPSID ASSEMBLY PROTEIN IS PHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21.
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X17403; CAA35353.1; -
 DR EMBL: X17403; CAA35354.1; ALT_INIT.
 DR FIR: S09843; Q08EB8.
 DR PDB: 1MPO; 15-OCT-97.
 DR PDB: 1LAY; 26-SEP-97.
 DR PDB: 1CMV; 04-SEP-97.
 DR PDB: 1IDA; 06-JUN-01.
 DR PDB: 1IEC; 06-JUN-01.
 DR PDB: 1IED; 06-JUN-01.
 DR PDB: 1IEF; 06-JUN-01.
 DR PDB: 1JG6; 06-JUN-01.
 DR PDB: 1JG6; 12-SEP-01.
 DR PDB: 1JG7; 12-SEP-01.
 DR PDB: 1NUT; 11-FEB-03.

DR PDB: 1NUT; 11-FEB-03.
 DR PDB: 1NRM; 11-FEB-03.
 DR PDB: 1NRM; 11-FEB-03.
 DR MEROPS: S21.002; -
 DR InterPro: IPR001847; Assemblin.
 DR Pfam: PF00716; Peptidase_S21.1.
 DR PRINTS: PR00236; HSCVCPSTDP40.
 KW Coat protein; Hydrolyase; Serine protease; Phosphorylation;
 KW 3D-structure.
 FT CHAIN 1
 FT PROPEP 257 643
 FT CHAIN 644 708
 FT CHAIN 1 708
 FT CHAIN 336 708
 FT CHAIN 393 708
 FT CHAIN 478 708
 FT SITE 256 257
 FT SITE 643 644
 FT ACT_SITE 63 63
 FT ACT_SITE 132 132
 FT ACT_SITE 157 157
 FT STRAND 14 23
 FT TURN 27 28
 FT TURN 30 31
 FT STRAND 34 34
 FT HELIX 36 44
 FT STRAND 58 61
 FT TURN 62 63
 FT STRAND 64 78
 FT TURN 79 80
 FT STRAND 81 88
 FT HELIX 91 101
 FT TURN 102 103
 FT HELIX 105 109
 FT TURN 113 114
 FT HELIX 119 127
 FT STRAND 130 133
 FT STRAND 158 161
 FT TURN 167 168
 FT STRAND 170 170
 FT STRAND 172 174
 FT HELIX 177 181
 FT TURN 182 183
 FT TURN 185 186
 FT HELIX 189 197
 FT TURN 198 199
 FT STRAND 214 214
 FT HELIX 218 229
 FT TURN 230 230
 FT TURN 232 233
 FT HELIX 246 246
 FT TURN 247 249
 FT HELIX 247 249
 SO SEQUENCE 708 AA; 73851 MW; 32A93D6586824C9 CRC64;
 Query Match 8.9%; Score 107.5; DB 1; Length 708;
 Best Local Similarity 27.1%; Pred. No. 4.1;
 Matches 69; Conservative 24; Mismatches 85; Indels 77; Gaps 14;
 QY 25 WLDPGHAAV-AQAPPVAVSS-SLEFDLSYTKLHHSLSQOSEPDLRH---LYLVVNTLRRIA 79
 DB 199 WQRCGSTAVDAGSDPFRSDSYGLGNSVDALY--IERLRPKLRDQVGTVERESYVA 256
 QY 80 SWAPAAAL-----PPV--PSP-----PAAPSV 99
 DB 257 SVSPEACDIKKAASERSGDSRSQAATTPAAGARVPSSSPVPEPSPVQPPALPRSPSV 316
 QY 100 ADNLASSDAALASMAWASLIEDLSHIEGLSQAPPL-----ADEGPPGRSIGGAAPSLGA 154
 DB 317 ---LPAESPPLSPSPAPAAASMSH-----PLSAVPAATAPATPGATVAGASPAVSS 364
 QY 155 LDLLGPATGCLLDD---GLEGLFEDIDTSMTDNELIMAPASGLKPGP-EDGPGKEADE 209

DB 365 --LAMPHGVLPKAFSLGASHANVPMTGCAVAPPSAPDLPLSTPASTYGAIV 422
 QY 210 IDEAL-----DYL 218
 DB 423 VGYDQLAARHPADYV 437

RESULT 9
 DUS8_MOUSE STANDARD; PRT; 663 AA.
 AC 009112;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dual specificity protein phosphatase 8 (EC 3.1.3.48) (EC 3.1.3.16)
 DE (Neuronal tyrosine threonine phosphatase 1).
 GN DUSP8 OR NTPPL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA MEDLINE=96311565; PubMed=8733337;
 RA Pateroson H., Mclellan-Arnold E., Boyd Y., Leversha M.A., Owen N.,
 RA Blake D.J., Ashworth A., Davies K.E.;
 RT "A member of the MAP kinase phosphatase gene family in mouse
 RT containing a complex trinucleotide repeat in the coding region.";
 RT Hum. Mol. Genet. 5:675-684(1996).
 CC -1- FUNCTION: THIS PROTEIN SHOWS BOTH ACTIVITY TOWARD TYROSINE-PROTEIN
 CC PHOSPHATE AS WELL AS WITH SERINE/THREONINE-PROTEIN PHOSPHATE (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
 CC tyrosine + phosphate.
 CC -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
 CC phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN AND LUNG.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY. DUAL SPECIFICITY SUBFAMILY.
 CC -1- SIMILARITY: Contains 1 rhodanese domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X95518; CAA64772.1; -;
 DR HSSP; Q16828; IMKP.
 DR MGD; MGI:106826; Dusp8.
 DR InterPro; IPR000340; DS_phosphatase.
 DR InterPro; IPR001763; Rhodanese-like.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR Pfam; PF00782; DSpC; 1.
 DR SMART; SM00195; DSpC; 1.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PS50206; RHODANES_3; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS0056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
 KW Hydroxylase; Nuclear protein.
 FT DOMAIN 23 138 RHODANES.
 FT 162 432 PROTEIN-TYROSINE PHOSPHATASE.
 FT 452 459 POLY-ARG.
 FT 555 558 POLY-SER.
 FT 559 576 POLY-GLY.
 FT 577 600 POLY-SER.
 FT DOMAIN 311 552 PRO-RICH.

FT ACT SITE 246 246 BY SIMILARITY.
 SQ SEQUENCE 663 AA: 68847 MW; 416F429A12C1FA7C CRC64;
 Query Match 8.8%; Score 106.5; DB 1; Length 663;
 Best Local Similarity 26.6%; Pred. No. 4.5;
 Matches 53; Conservative 21; Mismatches 66; Indels 59; Gaps 10;

QY 27 DPGHAANAQAPPAVASSSLFDLSVLKLNHSLOOSEPDRLHVLVYNT-----LRRIOASMA 82
 DB 270 DDATRFVADRRRSISPNFNFLGQLLEVERSLK-----LLAALQTGPHLGTPEPLMG 321
 QY 83 PAALPVPSPSPAASVADNLASDALSMSASLSDLSHIEGSAQPLADECPG 142
 DB 322 PAAG---IPLPRLPSTSESAATGSEAAATAAR-----EGSP- 354
 QY 143 RSIGGAAPSLGALDLGPATGCLDDGLEGFEDIDTSMYDNEI-----MAPAS 191
 DB 355 -SAGGDAP-----IPSTAPATSA-LQOGLRGHLHSSDRLODTNRKRFSLDIKSAVPSR 408
 QY 192 EGLKPPEDGPKKEAPEL 210
 DB 409 RPDFPGPD-PG--EAPKL 424

RESULT 10
 CNG4_HUMAN STANDARD; PRT; 909 AA.
 ID CNG4_HUMAN
 AC 014028; Q14029;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cyclic-nucleotide-gated cation channel 4 (CNG channel 4) (CNG-4)
 DE (CNG4) (Cyclic nucleotide-gated cation channel subunit).
 GN CNG4 OR CNGC4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS RCNC2A AND RCNC2B).
 RC TISSUE=Retina;
 RA MEDLINE=93226050; PubMed=7682292;
 RA Chen T.Y., Peng Y.-W., Dhallan R.S., Ahmed B., Reed R.R., Yau K.-W.;
 RT "A new subunit of the cyclic nucleotide-gated cation channel in
 RT retinal rods.";
 RL Nature 362:764-767(1993).
 CC -1- SUBUNIT: HETEROOLIGOMERIC COMPLEX WITH CNG1.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: Named isoforms-2;
 CC Event-Alternative splicing; Named isoforms-2;
 CC Name=RCNC2B;
 CC IsoId=Q14028-1; Sequence=Displayed;
 CC Name=RCNC2A;
 CC IsoId=Q14028-2; Sequence=VSP_001110;
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L15296; AAA65620.1; -;
 DR EMBL; L15297; AAA65619.1; -;
 DR MIM; 600724; -;
 DR GO; GO:0017071; C:intracellular cyclic nucleotide activated c. . .; NAS.
 DR GO; GO:0005222; F:intracellular cAMP activated cation channel. . .; NAS.
 DR InterPro; IPR000595; CNMP_binding.
 DR InterPro; IPR005821; Ion_trans.
 DR InterPro; IPR001622; K-channel_pore.

```

DR pfam: PF00027; CNMP_binding; 1.
DR pfam: PF00520; Ion_trans; 1.
DR SMART: SM00100; CNMP; 1.
DR PROSITE: PS00888; CNMP_BINDING_1; 1.
DR PROSITE: PS00889; CNMP_BINDING_2; 1.
DR PROSITE: PS50042; CNMP_BINDING_3; 1.
KW Ionic channel; Ion transport; CNMP-binding; Transmembrane;
KW Multigene family; Alternative splicing.
FT DOMAIN 1 314
FT TRANSMEM 315 333
FT DOMAIN 334 347
FT TRANSMEM 348 366
FT DOMAIN 367 391
FT TRANSMEM 392 411
FT DOMAIN 412 448
FT TRANSMEM 449 471
FT DOMAIN 472 515
FT TRANSMEM 516 535
FT DOMAIN 536 619
FT TRANSMEM 620 640
FT DOMAIN 641 909
FT NE_BIND 628 767
FT BINDING 688 688
FT BINDING 700 700
FT VARSPLIC 1 286
SQ SEQUENCE 909 AA: 102285 MW: DCOE754336BEDD CRC64;
Query Match 8.8%; Score 106.5; DB 1; Length 909;
Best Local Similarity 21.7%; Pred. No. 6.4;
Matches 51; Conservative 34; Mismatches 89; Indels 61; Gaps 10;
QY 11 EEEEREPLAVDSWMLPGHAAVAOAPVASSSLEDLVKLHSLQOSEPDLRLHLYV 70
Db 88 EEEAEKEP---QDM-----AEKKEPEAEAEKASGVATKQHPVEVEDPD----- 131
QY 71 VNTLRITQASMAPAALPPVPS---PPAAPVADNILLASDAALS--ASMASILEDLSHI 125
Db 132 -----ADSCPLAMENPSTVLPSPSPAKSDTLIVSSASGTHRKRLPSDEDAEEL 183
QY 126 EGLSOAPOPPLADGCP-----GRSIGGAAPSLGALDLGPRATGCLLDGLEGF---- 174
Db 184 KALSPAESPVAVASDFTTPKRDGQORAASTAS-----TNSAINDRLQELVKLER 234
QY 175 ---EDIDTSMYDNELM-----APASEGLKPGEDGKGEAPLDEAELDYL 218
Db 235 ERTEKKEKLLDPDVTSDSESPKSPAKKAPAPAPTRFAEAE-PVEEHYCDML 288
RESULT 11
ODP2_RHIME STANDARD; PRT: 447 AA.
AC 09R9N3;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase
DE complex (EC 2.3.1.12) (E2).
DE PDHC OR PDH OR R01447 OR SMC01032.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=20255051; PubMed=10796014;
RA Cabanes D., Boistard P., Batut J.;
RT "Symbiotic induction of pyruvate dehydrogenase genes from
RT Sinorhizobium meliloti.";
RL Mol. Plant Microbe Interact. 13:483-493(2000).
[2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaire V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallbert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -1- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
CC CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
CC COPIES OF THREE ENZYMAIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
CC DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
CC (E3) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide = CoA + S-
CC acetylhydrolipoamide.
CC -1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL
CC COFACTOR (BY SIMILARITY).
CC -1- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
CC SYMMETRY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
CC -1- SIMILARITY: Contains 1 lipoyl-binding domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF190792; AAF04589.1; -.
CC EMBL: AL591787; CAC46026.1; -.
DR HSSP: P07016; 1BBL.
DR InterPro: IPR001078; 2Oxoacid_dh.
DR InterPro: IPR006257; Acet_long.
DR InterPro: IPR000089; Acet_long.
DR InterPro: IPR004167; E3 binding.
DR InterPro: IPR003016; Lipoyl.
DR pfam: PF00198; 2-oxoacid_dh; 1.
DR pfam: PF00364; biotin_lipoyl; 1.
DR pfam: PF02817; e3_binding; 1.
DR ProDom: PD001115; 2Oxoacid_dh; 1.
DR TIGRFAMs: TIGR01349; PDHac_trf_mito; 1.
DR PROSITE: PS00189; LIPOYL; 1.
KW Glycolysis; Transferase; Acyltransferase; Lipoyl; Complete proteome.
FT BINDING 43 43 LIPOYL (BY SIMILARITY).
FT ACT_SITE 420 420 POTENTIAL.
SQ SEQUENCE 447 AA: 46140 MW: 48B1CAC4E2FDC2AC CRC64;
Query Match 8.7%; Score 105; DB 1; Length 447;
Best Local Similarity 31.1%; Pred. No. 3.7;
Matches 52; Conservative 14; Mismatches 51; Indels 50; Gaps 9;
QY 79 ASMAPAALPPVSPAPASVAD-----NLIASSDAALSMSASILEDLSHIGLSGAPR 134
Db 113 AAPAPAAPAPQAPAPASAPADGCKRTFSSPIARRLAKKEGI--DLAING----- 163
QY 135 LADEGPGRST-----GGAAPSLGALDLGAPATGCLLDGLEGLEDIDTSMYDNE 185
Db 164 ---SGPHGVYKVDVETAVSGAARPAKA-PAAPAPATL-----AKGNSDEAVLKF-- 213
QY 186 LMAPASEGLKPGPEDGPGKEAPLDEA-----ELDYLM 219
Db 214 ---PSGYELV--PHDGMRTTAKRLVESKOTIPHFYVSVCEDLALM 255
RESULT 12
ID FNM2_MOUSE
ID FNM2_MOUSE STANDARD; PRT: 1567 AA.
AC 09JL04;

```

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Formin 2.
 GN FMN2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain:
 RX MEDLINE=20245324; PubMed=10781961;
 RA Leader B., Leder P.;
 RT "Formin-2, a novel formin homology protein of the cappuccino
 RT subfamily, is highly expressed in the developing and adult central
 RT nervous system."
 RL Mech. Dev. 93:221-231(2000).
 CC -1- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN THE DEVELOPING
 CC AND MATURE CENTRAL NERVOUS SYSTEM.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION BEGINS AT EMBRYONIC DAY 9.5 IN
 CC THE DEVELOPING SPINAL CORD AND BRAIN STRUCTURES AND CONTINUES IN
 CC NEONATAL AND ADULT BRAIN STRUCTURES INCLUDING THE OLFACTORY BULB,
 CC COREX, THALAMUS, HYPOTHALAMUS, HIPPOCAMPUS AND CEREBELLUM.
 CC -1- SIMILARITY: Contains 1 Formin homology 1 (FH1) domain.
 CC -1- SIMILARITY: Contains 1 Formin homology 2 (FH2) domain.
 CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOGY FAMILY. CAPPUCCINO
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF218940; AAF72883.1; -
 DR MGD; MGI:1859252; Fmn2.
 DR InterPro; IPR003104; FH2.
 DR InterPro; IPR001265; Formin.
 DR Pfam; PF02181; FH2; 1.
 DR PRINTS; PRO0828; FORMIN.
 DR SMART; SM00498; FH2; 1.
 KM Developmental protein; Repeat; Coiled coil.
 FT DOMAIN 643 683
 FT FT 919 1039
 FT DOMAIN 919 1039
 FT REPEAT 919 929
 FT REPEAT 930 940
 FT REPEAT 941 951
 FT REPEAT 952 962
 FT REPEAT 963 973
 FT REPEAT 974 984
 FT REPEAT 985 995
 FT REPEAT 996 1006
 FT REPEAT 1007 1017
 FT REPEAT 1018 1028
 FT REPEAT 1029 1039
 FT REPEAT 1128 1532
 FT DOMAIN 1408 1444
 FT DOMAIN 48 55
 FT DOMAIN 202 207
 FT DOMAIN 797 801
 FT DOMAIN 861 864
 FT DOMAIN 908 917
 FT DOMAIN 922 928
 FT DOMAIN 933 939
 FT DOMAIN 944 950
 FT DOMAIN 955 961
 FT DOMAIN 966 972
 FT DOMAIN 977 983

FT DOMAIN 988 994 POLY-PRO.
 FT DOMAIN 999 1005 POLY-PRO.
 FT DOMAIN 1010 1016 POLY-PRO.
 FT DOMAIN 1021 1027 POLY-PRO.
 FT DOMAIN 1032 1038 POLY-PRO.
 FT DOMAIN 1043 1046 POLY-PRO.
 FT DOMAIN 1054 1060 POLY-PRO.
 FT DOMAIN 1065 1072 POLY-PRO.
 FT DOMAIN 1077 1080 POLY-PRO.
 SQ SEQUENCE 1567 AA; 166268 MW; 8F273B1C8505944 CRC64;
 Query Match 8.6%; Score 103.5; DB 1; Length 1567;
 Best Local Similarity 20.4%; Pred. No. 18;
 Matches 64; Conservative 33; Mismatches 100; Indels 117; Gaps 12;
 QY 4 GGLKRRREEEKEPLAVDSWMLPGHAANVQAP----- 37
 DB 84 KGACDSRDVADSQLPTGE--LDSAHIVYKRTDLSABETGLSDTECADPREVTHPG 141
 QY 38 ---PAVA-----SSLPDLVYKLIHSLQOSE--PDLRHLYLV 70
 DB 142 ASRPAAAGVGIQATAEDELTAAGAAGDGGRTSSGSDTDIYSFHSATGEODLSDIQAIRL 201
 QY 71 VNTLRR---IQSNAPAAALPPV--PSPAPASVDNLASSDA-----L 111
 DB 202 QQQQQQKLLLDSEPPAA--PPTAISPPQAFGLGDPLGPRSEAEKDVQALPVRPDL 259
 QY 112 SASNASLLED-----LSHIEGLSQAPPLADE--GPPGRSIGAPASIGALDILGPA 161
 DB 260 PETKRSIVPEHPSPSSGSHLTSETGYPATAPASAVTDSLSPPAFPTPEAGPGGAGVPPAG 319
 QY 162 TGCLLDGLBGLFEDIDTSMTDNEIMAPASGLKRPEDGCKEAPELDAELDYLMGV 221
 DB 320 TGDDEEECEDEAFED-----APRGSPEGEWPEMEBA----- 351
 QY 222 LVGTQALERRPPGP 235
 DB 352 ---SORLKEPEEG 362
 RESULT 13
 ID A180_MOUSE STANDARD; PRT: 901 AA.
 AC 061548; 061547;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Clathrin coat assembly protein APl80 (Clathrin coat associated protein
 DE APl80) (91 kDa synaptosomal-associated protein) (Phosphoprotein Fl-
 DE 20).
 GN SNAP91.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RX MEDLINE=92300439; PubMed=1607933;
 RA Zhou S., Sousa R., Tannery N.H., Lafer E.M.;
 RT "Characterization of a novel synapse-specific protein. II. cDNA
 RT cloning and sequence analysis of the Fl-20 protein."
 RL J. Neurosci. 12:2144-2155(1992).
 CC -1- FUNCTION: ADAPTINS ARE COMPONENTS OF THE ADAPTOR COMPLEXES WHICH
 CC LINK CLATHRIN TO RECEPTORS IN COATED VESICLES. CLATHRIN-ASSOCIATED
 CC PROTEIN COMPLEXES ARE BELIEVED TO INTERACT WITH THE CYTOPLASMIC
 CC TAILS OF MEMBRANE PROTEINS, LEADING TO THEIR SELECTION AND
 CC CONCENTRATION. BINDING OF APl80 TO CLATHRIN TRISKELIA INDUCES
 CC THEIR ASSEMBLY INTO 60-70 NM COATS.
 CC -1- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE
 CC CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=long;

```

CC      IsoId=061548-1; Sequence=Displayed;
CC      Name=Short;
CC      IsoId=061548-2; Sequence=VSP_000172;
CC      TISSUE SPECIFICITY: BRAIN. ASSOCIATED WITH THE SYNAPSES.
CC      -1- DEVELOPMENTAL STAGE: DEVELOPMENTALLY REGULATED IN A PATTERN
CC      COINCIDENT WITH ACTIVE SYNAPTOGENESIS AND SYNAPTIC MATURATION.
CC      -1- DOMAIN: POSSESSES A THREE DOMAIN STRUCTURE: THE N-TERMINAL 300
CC      RESIDUES HARBOR A CLATHRIN BINDING SITE, AN ACIDIC MIDDLE DOMAIN
CC      450 RESIDUES, INTERRUPTED BY AN ALA-RICH SEGMENT, AND THE C-
CC      TERMINAL DOMAIN (166 RESIDUES).
CC      -1- PTM: Phosphorylated.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC      or send an email to license@isb-slb.ch).
CC      -----
CC      EMBL; M83985; AAA37587.1; -.
CC      EMBL; M83985; AAA37586.1; -.
CC      PIR; A44825; A44825.
CC      MGD; MGI:109132; Snapp1.
CC      InterPro: IPR001026; ENTH.
CC      Pfam; PF01417; ENTH; 1.
CC      SMART; SM00273; ENTH; 1.
CC      Coated pits; Alternative splicing; Phosphorylation.
CC      FT DOMAIN 410 413 POLY-THR.
CC      FT DOMAIN 535 539 POLY-ALA.
CC      FT DOMAIN 547 550 POLY-ALA.
CC      FT DOMAIN 659 664 POLY-SER.
CC      FT DOMAIN 704 710 POLY-SER.
CC      VARSPLIC 715 719 Missing (In isoform short).
CC      FT FT /FtId=VSP_000172.
CC      SQ SEQUENCE 901 AA; 91851 MW; 24A98FBACEB8B1 CRC64;
CC
CC      Query Match 8.5%; Score 102.5; DB 1; Length 901;
CC      Best Local Similarity 21.2%; Pred. No. 12;
CC      Matches 53; Conservative 25; Mismatches 93; Indels 79; Gaps 9;
CC
CC      QY 32 AQAQAPPAVASSSLEFLDSYLKLHHSIQSEPDRLHVLVYNTLRIRIQASMAPAALPP-- 89
CC      Db 518 APSAPPAVAVAAATTAATTTA-----AAAATTTATTSAAATTTAAPAL 559
CC      QY 90 -----VSPPAAPASV---ADNLASSDPAALSASMASLLEDLSHIEGLSQ 130
CC      Db 560 DIFGDLFDPSAEVAAAPKPDPAAPSIDLFGTDAFSPSPGASPVPESSITADLLSVDATA- 618
CC      QY 131 APOPLADGCPGKRSIGCAAPSGLGALDLGPATGCLLDLGLGLEPD-----IDTSM-- 181
CC      Db 619 APSAPSTASP-----AKAESSGVIDLDFGDAFG-----SGASETPAPAPQAVSSSSASA 665
CC      QY 182 -----YNEELMAPASEGLKPPED-----GPKGEAPLEDAEALDYLMDVYVG 224
CC      Db 666 DLAGEFGSEMAPFTPTVPYQNNLLDPSEFAAFCTPTSTSSSSSFPDSEVDGLGLLMP 725
CC      QY 225 TQALEPPGP 234
CC      Db 726 TMAPSGQAP 735
CC
CC      RESULT 14
CC      SRCA_RABIT STANDARD; PRT; 908 AA.
CC      AC P13666;
CC      DT 01-JAN-1990 (Rel. 13, Created)
CC      DT 01-JAN-1990 (Rel. 13, Last sequence update)
CC      DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC      DE Sarcalamenia precursor.
CC      OS Sarcotylagus cuniculus (Rabbit).
CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

```

```

XX NCBI_TaxID=9986;
RN
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=89345602; PubMed=2762314;
RA Leberer E., Charuk J.H.M., Green N.M., MacLennan D.H.:
RT "Molecular cloning and expression of cDNA encoding a luminal calcium binding glycoprotein from sarcoplasmic reticulum.",
RL Proc. Natl. Acad. Sci. U.S.A. 86:6047-6051(1989).
[2]
RN
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=89123480; PubMed=2521635;
RA Leberer E., Charuk J.H.M., Clarke D.M., Green N.M., Zubrycka-Gaarn E.,
McLennan D.H.:
RT "Molecular cloning and expression of cDNA encoding the 53,000-dalton glycoprotein of rabbit skeletal muscle sarcoplasmic reticulum.",
RL J. Biol. Chem. 264:3484-3493(1989).
CC -1- FUNCTION: PERHAPS INVOLVED IN THE REGULATION OF CALCIUM TRANSPORT.
CC -1- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN. ASSOCIATED THROUGH CA(2+) WITH THE MEMBRANE.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named Isoforms=2;
CC Name=1; Synonyms=160 KDa;
CC IsoId=P13666-1; Sequence=Dsplayed;
CC Name=?; Synonyms=58 kDa;
CC IsoId=P13666-2; Sequence=VSP_004429;
CC Note=No experimental confirmation available;
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/a-slb.ch).
CC or send an email to license@isb-slb.ch).
DR
EMBL: M25750; AAA31189.1; -.
DR EMBL: J04480; AAA60730.1; -.
DR PIR: A33280; A33280.
KW Calcium-binding; Glycoprotein; signal; Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 908 SARCALOUMENIN.
FT DOMAIN 20 457 ACIDIC DOMAIN, PROBABLY BINDS CALCIUM.
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 716 716 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 20 455 Missing (in isoform 2).
FT FT /FTid=VSP_004429.
SQ CONFLICT 474 474 Q->E (IN REF. 2).
SQ SEQUENCE 908 AA; 97920 MW; A48CAA221AE141B8 CRC64;

Query Match 8.5%; Score 102.5; DB 1; Length 908;
Best Local Similarity 26.2%; Pred. NO. 12;
Matches 63; Conservative 24; Mismatches 104; Indels 49; Gaps 11

QY 32 AVAQAAPPAVAASSLFELSVLKLAH-----SLQOSE-----PDLRHLVLVNTLTTRI 77
DY 19 AELQVP---ASGGETEDVCNLEHNHFSSAGASILEKERATLYADAAPRDENLLHYDGREA 75
QY 78 QASNAAPAAALPP-VPSPPAASVDNLNLIASSDAISA-----SMASLEDL 122
DY 76 ESARTTAGAPPAPAPGPDSPEAS--LPNNSATESAPGPGDATGPREROGAPAAASALPPG 132
QY 123 SHIEGLSOAPOPLAD-RGGPERSIGGAAPSGLG--DLISPATGCLLDLGLEGLEFIDT 179
DY 133 GDESSLDEROELSSEGGEGEAAGLGLIPSGAASGGKGOAGGKVKREAGEVGADSPV 192
QY 180 SMDYNELMAPASEGLDPGED-----GPCK-EAAPDLDAEILDYLMDEVLTQALRRP 231
DY 193 QGAALAEPAEPASIASSEDQHTLEGEKKGSPGDHGFTELDGFDD--GASAGEEP 249

```



```
FT DOMAIN 962 1034 TRIPLE-HELICAL REGION 6 (COL6).
FT DOMAIN 1035 1044 NONHELICAL REGION 7 (NC7).
FT DOMAIN 1045 1077 TRIPLE-HELICAL REGION 7 (COL7).
FT DOMAIN 1078 1089 NONHELICAL REGION 8 (NC8).
FT DOMAIN 1090 1111 TRIPLE-HELICAL REGION 8 (COL8).
FT DOMAIN 1112 1118 NONHELICAL REGION 9 (NC9).
FT DOMAIN 1119 1173 TRIPLE-HELICAL REGION 9 (COL9).
FT DOMAIN 1174 1186 NONHELICAL REGION 10 (NC10).
FT DOMAIN 1187 1204 TRIPLE-HELICAL REGION 10 (COL10).
FT DOMAIN 1205 1516 NONHELICAL REGION 11 (NC11).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 691 691 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1329 1329 O-LINKED (GALNAC. . .) (POTENTIAL).
FT DISULFID 1366 1506 BY SIMILARITY.
FT DISULFID 1468 1498 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1095 1097 MISSING (in Isoform Short).
FT VARSPLIC 1 180
FT VARSPLIC 181 215
FT VARSPLIC 1437 1437
FT VARIANT 1437 1437
FT CONFLICT 428 428 /FTid-VSP_001156.
FT CONFLICT 841 841 D -> N (increased risk of developing
FT CONFLICT 877 877 prostate cancer).
FT CONFLICT 886 886 /FTid-VAR_012709.
FT CONFLICT 912 912 F -> S (IN REF. 2).
FT CONFLICT 933 933 I -> V (IN REF. 2).
FT CONFLICT 975 975 V -> L (IN REF. 3).
FT CONFLICT 1064 1064 P -> R (IN REF. 3).
FT CONFLICT 1084 1084 R -> L (IN REF. 3).
FT CONFLICT 1120 1120 P -> L (IN REF. 3).
FT CONFLICT 1123 1123 A -> P (IN REF. 3).
FT CONFLICT 1126 1126 L -> K (IN REF. 3).
FT CONFLICT 1206 1206 P -> A (IN REF. 3).
FT CONFLICT 1304 1304 P -> A (IN REF. 3).
FT CONFLICT 1314 1314 P -> A (IN REF. 3).
FT CONFLICT 1323 1324 G -> GQ (IN REF. 3).
FT CONFLICT 1443 1443 R -> G (IN REF. 3).
FT CONFLICT 1483 1483 R -> CG (IN REF. 3).
FT CONFLICT 1516 1516 LR -> CG (IN REF. 3).
FT CONFLICT 153840 153840 R -> T (IN REF. 4).
FT CONFLICT 1483 1483 S -> Y (IN REF. 4).
SO SEQUENCE 1516 AA; 153840 MM; 3C70E29A4476EE76 CRC64;

Query Match 8.4%; Score 102; DB 1; Length 1516;
Best Local Similarity 27.3%; Pred. No. 22;
Matches 47; Conservative 9; Mismatches 58; Indels 58; Gaps 9;

QY 77 IQASMAPAAALPPVSPPAAPSVADNL---LASSDAALASMASL--LEDLSHTEGLSQ 130
Db 660 LRLALNGP-RGPPGPPGPGVPGIPEGPFYGVNSSDVGPAIGVPGREGPPGPPGLPG 718
QY 131 APOPLADEGPPGR---SIG--GAAPSGALDLGLPATGCLLDGLEGLEFEDIDTSMTD 183
Db 719 PPGPPGREGPPGPPRTGKGSLGEGAGPHGKSGKAGPRA-GARGESGLAG----- 766
QY 184 NEIMAPASBGLKPPGPDGPKGEAPLDEALDYLMVDLVVQALERRPPPG 235
Db 767 ---APGPAG-PPGPPGPPG-----PPGP 786
```

Search completed: August 19, 2003, 09:33:10
Job time : 15 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:31:59 ; Search time 97 Seconds
(without alignments)
627,839 Million cell updates/sec

Title: US-09-701-675A-3

Perfect score: 1210
Sequence: 1 MUSKGLKRRKEEKEKPLA.....YLMQVLTGQALRPPGGR 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTRMBL_23: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_prodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriophage: *
17: sp_archaeal: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 1210 | 100.0 | 236 | 4 | Q9UHV2 |
| 2 | 1206 | 99.7 | 236 | 4 | Q9BUE7 |
| 3 | 1044 | 86.3 | 236 | 11 | Q9ULI0 |
| 4 | 1032 | 85.3 | 236 | 11 | Q9D888 |
| 5 | 968 | 80.0 | 223 | 11 | Q925E6 |
| 6 | 792 | 65.5 | 244 | 11 | Q9DCZ2 |
| 7 | 215 | 17.8 | 309 | 11 | Q91WL3 |
| 8 | 215 | 17.8 | 309 | 11 | Q91TGS |
| 9 | 215 | 17.8 | 309 | 11 | Q91JGS |
| 10 | 214 | 17.7 | 197 | 11 | Q9EC3 |
| 11 | 213 | 17.6 | 196 | 4 | Q96CQ2 |
| 12 | 213 | 17.6 | 196 | 4 | Q9UJW9 |
| 13 | 170 | 14.0 | 241 | 4 | Q8TP18 |
| 14 | 165 | 13.6 | 241 | 4 | Q9BXL8 |
| 15 | 152 | 12.6 | 237 | 11 | Q9CWM2 |
| 16 | 134 | 11.1 | 207 | 4 | Q9NWK7 |

| | | | | | | |
|----|-------|------|------|----|--------|---------------------|
| 17 | 125.5 | 10.4 | 181 | 11 | Q925E5 | Q925E5 mus musculus |
| 18 | 121.5 | 10.0 | 1339 | 11 | Q35788 | Q35788 rat norv |
| 19 | 120.5 | 10.0 | 752 | 10 | Q8LK21 | Q8LK21 chlamydom |
| 20 | 120 | 9.9 | 3164 | 12 | Q69088 | Q69088 human herpe |
| 21 | 117 | 9.7 | 446 | 11 | Q8YD23 | Q8YD23 mus musculus |
| 22 | 115.5 | 9.5 | 840 | 4 | Q8N3F8 | Q8N3F8 homo sapien |
| 23 | 115 | 9.5 | 698 | 10 | Q9FE49 | Q9FE49 chlamydom |
| 24 | 115 | 9.5 | 699 | 10 | Q9FED4 | Q9FED4 chlamydom |
| 25 | 114 | 9.4 | 336 | 4 | Q8IXT4 | Q8IXT4 homo sapien |
| 26 | 114 | 9.4 | 1313 | 5 | Q9NE28 | Q9NE28 leishmania |
| 27 | 114 | 9.4 | 2137 | 4 | Q15021 | Q15021 homo sapien |
| 28 | 113 | 9.3 | 963 | 11 | Q9QXMT | Q9QXMT mus musculus |
| 29 | 112 | 9.3 | 715 | 4 | Q9H0L5 | Q9H0L5 homo sapien |
| 30 | 110 | 9.3 | 791 | 4 | Q9BY92 | Q9BY92 homo sapien |
| 31 | 110 | 9.1 | 202 | 11 | Q921E8 | Q921E8 homo sapien |
| 32 | 109 | 9.0 | 657 | 4 | Q8N3N8 | Q8N3N8 homo sapien |
| 33 | 109 | 9.0 | 688 | 12 | Q65575 | Q65575 bovine herp |
| 34 | 108.5 | 9.0 | 1531 | 4 | Q60336 | Q60336 bovine herp |
| 35 | 108 | 8.9 | 941 | 16 | Q9Z1W5 | Q9Z1W5 streptomyce |
| 36 | 107 | 8.8 | 993 | 11 | Q8CHE2 | Q8CHE2 mus musculus |
| 37 | 107 | 8.8 | 1317 | 11 | Q35211 | Q35211 mus musculus |
| 38 | 107 | 8.8 | 1317 | 11 | Q8CB63 | Q8CB63 mus musculus |
| 39 | 106.5 | 8.8 | 512 | 17 | Q9HPL5 | Q9HPL5 halobacteri |
| 40 | 106.5 | 8.8 | 753 | 16 | Q87124 | Q87124 pseudomonas |
| 41 | 106.5 | 8.8 | 1245 | 4 | Q43636 | Q43636 homo sapien |
| 42 | 106.5 | 8.8 | 1251 | 4 | Q9UMG2 | Q9UMG2 burkholderi |
| 43 | 106 | 8.8 | 442 | 9 | Q8HAP5 | Q8HAP5 burkholderi |
| 44 | 106 | 8.8 | 993 | 4 | Q9HC9 | Q9HC9 homo sapien |
| 45 | 104 | 8.6 | 539 | 11 | Q8CFW3 | Q8CFW3 mus musculus |

ALIGNMENTS

RESULT 1

Q9UHV2 PRELIMINARY; PRT; 236 AA.

AC Q9UHV2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE CDK4-binding protein p34SEI1 (TRIP-Brl) (SCDK4-binding protein p34SEI1).
 GN SEI1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=20047903; PubMed=10580009;
 RA Sugimoto M., Nakamura T., Ohtani N., Hampson I.N.,
 RA Shimoto A., Furuchi Y., Okumura K., Niva S., Taya Y., Hara E.,
 RT "Regulation of CDK4 activity by a novel CDK4 binding protein,
 RT p34SEI1.";
 RL Genes Dev. 13:3027-3033(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21231173; PubMed=11331592;
 RA Hsu S.T., Yang C.M., Sim K.G., Hentschel D.M., O'Leary E.,
 RA Bonventre J.V.,
 RT "TRIP-Brl: a novel family of PHD zinc finger- and bromodomain-
 RT interacting proteins that regulate the transcriptional activity of
 RT E2F-1/DP-1.";
 RL EMBO J. 20:2273-2285(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Rieder M.J., Braun A.C., Montoya M.A., Chung M.-W., Nguyen C.P.,
 RA Nguyen D.A., Livingston R.J., Poel C.L., Robertson P.D.,
 RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.,
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBD databases.
 DR EMBL; AF117959; AAF08349.1; -;
 EMBL; AF366402; AAK52831.1; -;

DR EMBL: AY130860; AAM77800.1; -
SQ SEQUENCE 236 AA; 24673 MW; 92F25EA328F155B1 CRC64;

Query Match 100.0%; Score 1210; DB 4; Length 236;
Best Local Similarity 100.0%; Pred. No. 7.9e-86;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAAPPVASSSLFDLSVTKLHHSLOOS 60
DB 1 MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAAPPVASSSLFDLSVTKLHHSLOOS 60
OY 61 EPDLRHLYLVNTLRRIQASMAPAALPPVSPPAASVADNLLASSDAALSASMASLLE 120
DB 61 EPDLRHLYLVNTLRRIQASMAPAALPPVSPPAASVADNLLASSDAALSASMASLLE 120
OY 121 DLSHTEGISOAPQPLADGPPGRSISGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
DB 121 DLSHTEGISOAPQPLADGPPGRSISGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
OY 181 MYDNELMAPASEGLKPGPEDGPGKEAPDELDEALDYLMVDLVGTQALERPPGGR 236
DB 181 MYDNELMAPASEGLKPGPEDGPGKEAPDELDEALDYLMVDLVGTQALERPPGGR 236

RESULT 2

O9BUE7 PRELIMINARY; PRT; 236 AA.

DR 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE CDK-binding protein p34Seil.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
DR Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC002670; AA02670.1; -
SQ SEQUENCE 236 AA; 24703 MW; 8659505783FBF50B CRC64;

Query Match 99.7%; Score 1206; DB 4; Length 236;
Best Local Similarity 99.6%; Pred. No. 1.6e-85;
Matches 235; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAAPPVASSSLFDLSVTKLHHSLOOS 60
DB 1 MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAAPPVASSSLFDLSVTKLHHSLOOS 60
OY 61 EPDLRHLYLVNTLRRIQASMAPAALPPVSPPAASVADNLLASSDAALSASMASLLE 120
DB 61 EPDLRHLYLVNTLRRIQASMAPAALPPVSPPAASVADNLLASSDAALSASMASLLE 120
OY 121 DLSHTEGISOAPQPLADGPPGRSISGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
DB 121 DLSHTEGISOAPQPLADGPPGRSISGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
OY 181 MYDNELMAPASEGLKPGPEDGPGKEAPDELDEALDYLMVDLVGTQALERPPGGR 236
DB 181 MYDNELMAPASEGLKPGPEDGPGKEAPDELDEALDYLMVDLVGTQALERPPGGR 236

RESULT 3

O9JUL10 PRELIMINARY; PRT; 236 AA.

DR 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE P34SEI-1 (1110032C13RIK protein) (TRIP-BR1) (RIKEN CDNA 1110032C13

DE (gene).
GN SEI1 OR 1110032C13RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;

RN [1]
RP SEQUENCE FROM N.A.
RA Ohtani N., Hara E.;
RT Cloning of mouse SEI-1 cDNA.*
RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shigaawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glass C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stanbly F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlay M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombert P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Winding L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.*";
RL Nature 409:685-690(2001).

RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21231173; PubMed=11331592;
RA Hsu S.I., Yang C.M., Sim K.G., Hentschel D.M., O'Leary E.,
RA Bonventre J.V.;
RT "TRIP-BR: a novel family of PHD zinc finger- and bromodomain-
interacting proteins that regulate the transcriptional activity of
E2F-1/DP-1.";
RL EMBO J. 20:2273-2285(2001).

RN [4]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF218291; AAF37653.1; -
DR EMBL: AK004022; BAB23130.1; -
DR EMBL: AF366401; AAK52830.1; -
DR EMBL: BC016077; AAH16077.1; -
DR MGI:1913438; 1110032C13RIK.
SQ SEQUENCE 236 AA; 25136 MW; 989ADFE29D84C5 CRC64;

Query Match 86.3%; Score 1044; DB 11; Length 236;
Best Local Similarity 86.4%; Pred. No. 5.2e-73;
Matches 204; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

OY 1 MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAAPPVASSSLFDLSVTKLHHSLOOS 60
DB 1 MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAAPPVASSSLFDLSVTKLHHSLOOS 60
OY 61 EPDLRHLYLVNTLRRIQASMAPAALPPVSPPAASVADNLLASSDAALSASMASLLE 120
DB 61 EPDLRHLYLVNTLRRIQASMAPAALPPVSPPAASVADNLLASSDAALSASMASLLE 120
OY 121 DLSHTEGISOAPQPLADGPPGRSISGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
DB 121 DLSHTEGISOAPQPLADGPPGRSISGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
OY 181 MYDNELMAPASEGLKPGPEDGPGKEAPDELDEALDYLMVDLVGTQALERPPGGR 236
DB 181 MYDNELMAPASEGLKPGPEDGPGKEAPDELDEALDYLMVDLVGTQALERPPGGR 236

```

Db      181 MYDSELMLPASBGLKPGPNKPAKEBPPELDEALDYLDMDVLTGTOALERPPEGGR 236

RESULT 4
09D888      PRELIMINARY;      PRT;      236 AA.

AC      09D888;
DT      01-JUN-2001 (TREMBlrel. 17, Created)
DT      01-JUN-2001 (TREMBlrel. 17, last sequence update)
DE      1110032C13Rik protein.
GN      1110032C13Rik.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57Bl/6J; TISSUE=Small Intestine;
RX      MEDLINE=21085660; PubMed=11217851;
RA      Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA      Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA      Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA      Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA      Fletschmann W., Gaasterland T., Gissi C., Kling B., Kochiwa H.,
RA      Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA      Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA      Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA      Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA      Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA      Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA      Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momhaerts P.,
RA      Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA      Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA      Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA      Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA      Hayashizaki Y.;
RT      "Functional annotation of a full-length mouse cDNA collection.";
RL      Nature 409:685-690(2001).
DR      EMBL; AK008303; BAB25588.1;
DR      MGD; MGI:1913438; 1110032C13Rik.
SQ      SEQUENCE 236 AA; 25023 MW; 95AE0BCFACA77549 CRC64;

Query Match      85.3%; Score 1032; DB 11; Length 236;
Best Local Similarity 85.6%; Pred. No. 4.4e-72;
Matches 202; Conservative 10; Mismatches 24; Indels 0; Gaps 0;

QY      1 MTSKGLKRRREEREEKPLAVDSWMLDPGHAAYQAQPPAVASSSLFDLSVTKLHSLQGS 60
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1 MTSKGLKRRREEREEKPLAVDSWMLDPGHAAYQAQPPAVASSSLFDLSVTKLHSLQGS 60
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      61 EPDLRLVLVYVNTLRRIQASMAAPALPPVSPAPASVADNLLASSDAALSMSASLLE 120
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      61 EPDLRLVLVYVNTLRRIQASMAAPALPPVSPAPASVADNLLASSDAALSMSASLLE 120
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      121 DLSHIEGLSQAPDLADEGPPGRSIGGAPASLGALDLPATGCLLDGEGLEGFEDIDTS 180
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      121 DLSHIEGLSQAPDLADEGPPGRSIGGAPASLGALDLPATGCLLDGEGLEGFEDIDTS 180
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      181 MYDNEIMAPASBGLKPGPNKPAKEBPPELDEALDYLDMDVLTGTOALERPPEGGR 236
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      181 MYDNEIMAPASBGLKPGPNKPAKEBPPELDEALDYLDMDVLTGTOALERPPEGGR 236
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
09D5E6      PRELIMINARY;      PRT;      223 AA.

AC      09D5E6;
DT      01-DEC-2001 (TREMBlrel. 19, Created)
DT      01-DEC-2001 (TREMBlrel. 19, last sequence update)
DE      1110032C13Rik protein.
GN      1110032C13Rik.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57Bl/6J; TISSUE=Small Intestine;
RX      MEDLINE=21085660; PubMed=11217851;
RA      Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA      Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA      Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA      Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA      Fletschmann W., Gaasterland T., Gissi C., Kling B., Kochiwa H.,
RA      Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA      Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA      Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA      Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA      Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA      Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA      Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momhaerts P.,
RA      Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA      Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA      Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

```

```

GN      1110032C13Rik.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57Bl/6J; TISSUE=Small Intestine;
RX      MEDLINE=21085660; PubMed=11217851;
RA      Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA      Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA      Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA      Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA      Fletschmann W., Gaasterland T., Gissi C., Kling B., Kochiwa H.,
RA      Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA      Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA      Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA      Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA      Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA      Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA      Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momhaerts P.,
RA      Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA      Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA      Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

Query Match      80.0%; Score 968; DB 11; Length 223;
Best Local Similarity 85.2%; Pred. No. 3.6e-67;
Matches 190; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

QY      1 MTSKGLKRRREEREEKPLAVDSWMLDPGHAAYQAQPPAVASSSLFDLSVTKLHSLQGS 60
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1 MTSKGLKRRREEREEKPLAVDSWMLDPGHAAYQAQPPAVASSSLFDLSVTKLHSLQGS 60
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      61 EPDLRLVLVYVNTLRRIQASMAAPALPPVSPAPASVADNLLASSDAALSMSASLLE 120
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      61 EPDLRLVLVYVNTLRRIQASMAAPALPPVSPAPASVADNLLASSDAALSMSASLLE 120
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      121 DLSHIEGLSQAPDLADEGPPGRSIGGAPASLGALDLPATGCLLDGEGLEGFEDIDTS 180
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      121 DLSHIEGLSQAPDLADEGPPGRSIGGAPASLGALDLPATGCLLDGEGLEGFEDIDTS 180
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      181 MYDNEIMAPASBGLKPGPNKPAKEBPPELDEALDYLDMDVLTGTOALERPPEGGR 236
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      181 MYDNEIMAPASBGLKPGPNKPAKEBPPELDEALDYLDMDVLTGTOALERPPEGGR 236
        ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
09DC22      PRELIMINARY;      PRT;      244 AA.

AC      09DC22;
DT      01-JUN-2001 (TREMBlrel. 17, Created)
DT      01-JUN-2001 (TREMBlrel. 17, last sequence update)
DE      1110032C13Rik protein.
GN      1110032C13Rik.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57Bl/6J; TISSUE=Small Intestine;
RX      MEDLINE=21085660; PubMed=11217851;
RA      Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA      Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA      Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA      Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA      Fletschmann W., Gaasterland T., Gissi C., Kling B., Kochiwa H.,
RA      Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA      Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA      Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA      Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA      Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA      Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA      Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momhaerts P.,
RA      Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA      Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA      Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

```


OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_Taxid=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=2354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL: AK076787; BAC36480.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 309 AA; 33312 MW; D4178688F0DF8F00 CRC64;
 Query Match 17.8%; Score 215; DB 11; Length 309;
 Best Local Similarity 27.8%; Pred. No. 7, 9e-09;
 Matches 90; Conservative 33; Mismatches 87; Indels 114; Gaps 12;
 QY 1 MSLGKRRREEREEKEPLAVDSWMLDPGHAAPPAVA---SSSLFDSLVLKHS 56
 DB 1 MSLGKRRREEREEKEPLAVDSWMLDPGHAAPPAVA---SSSLFDSLVLKHS 51
 QY 57 LQSEPDRLHVLVYVNTLRRIQ-----ASMAPA-----AALPPVSP- 93
 DB 52 RLPTSPLOKTVLINMLRRIQEEELKQEGSLRPAFTSSQPSNSISDSYQEAAPPAPHP 111
 QY 94 -----PAAPSYADNLLASDAALASMSALILE 120
 DB 112 DGGSTTPLEACLPASLEDDNDTRCTIQAVHPAAPT-----RLSSAALRAEKDSESS 165
 QY 121 DLSHTEGL-----SOAPQPLADEGPPGRSIGA-----APSLGALD-----L 157
 DB 166 ALDEIEELCPTSTSTGAAHTAPEGKGTSSSSVOKPGPREGRTDGRFMDSLPGNFE 225
 QY 158 LGPANGCCLLDGLEG-LFEDIDTSMYDNLNAPAS---EGLAPRGDGGKEEAP----- 208
 DB 226 ITTSTGFTLDTLDFADIDTSMYDFDPCSTGASGTASKMAVSDLLKTLAPYSNOP 285
 QY 209 -----ELDEAELDYIMDYVGT 225
 DB 286 VAPSQPFKMDLDELHIMEVLVGS 309
 RESULT 10
 Q9ERC3 PRELIMINARY; PRT; 197 AA.
 AC Q9ERC3:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DE 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Replication protein-binding trans-activator RBR1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_Taxid=10090;
 RP SEQUENCE FROM N.A.
 RC Cho J.M., Alaoui-Jamali M.A.;
 RT "Cloning of mouse RBR1 cDNA."
 RL Submitted (OCR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Tissue=Mammary gland;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF317202; AAG30951.1; -
 DR EMBL: BC034886; AAH34886.1; -
 DR InterPro: IPR000194; ATPase_a/centre.
 DR PROSITE: PS00152; ATPase_ALPHA_BETA; 1.
 SQ SEQUENCE 197 AA; 21934 MW; CFF6FAAC0A91D7E7 CRC64;
 Query Match 17.7%; Score 214; DB 11; Length 197;

Best Local Similarity 31.8%; Pred. No. 5, 5e-09;
 Matches 74; Conservative 37; Mismatches 72; Indels 50; Gaps 13;
 QY 5 GLKRRKREEREEKEPLAVDSW-MLDPGHAAPPAVAS--SSLPDSLVLKHSIQ 58
 DB 3 GLKRRKSDLEEEEEE-----KWDW-----SPTLRSTQOALLRISLDKVRSLG 47
 QY 59 QSEPDRLHVLVYVNTLRRIQAS--MAPAALPPVSPPAAPSVADNLLASDAALASMA 116
 DB 48 PRAPSLRRVHLINHTLQIQALRLAPAPALPPEPL-----FLGEEFSLSTTIG 97
 QY 117 SLLEDL-SHIEGLSQAPPLADEGPPGRSIGAAPSLGALLDLPATGCLLDGEGLE 175
 DB 98 SLIRELDTSMDEEPLNPAASSPONEIATVQADVPF--LEAL--SSRYLGDSGLDDEFL 153
 QY 176 DIDTSMYDNLNAPASEGLKPGED--GPGKEAPELDEALDYIMDYVGT 225
 DB 154 DIDTSAVERKVALPPE---PPHSIFCSGSE---WNELDHIMEITIGS 197
 RESULT 11
 Q96C02 PRELIMINARY; PRT; 196 AA.
 AC Q96C02:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_Taxid=9606;
 RP SEQUENCE FROM N.A.
 RC Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC014061; AAH14061.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 196 AA; 21768 MW; 9C54FAFA126F17AF1 CRC64;
 Query Match 17.6%; Score 213; DB 4; Length 196;
 Best Local Similarity 32.6%; Pred. No. 6, 5e-09;
 Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;
 QY 5 GLKRRKREEREEKEPLAVDSW-MLDPGHAAPPAVAS--SSLPDSLVLKHSIQSEPD 63
 DB 4 GLKRRKSDLEEEEEE-----ERWESPAGLSYQQA-----LRLSLDRVQSLGPRAPS 51
 QY 64 LRLHVLVYVNTLRRIQAS--MAPAALPPVSPPAAPSVADNLLASDAALASMSALIED 121
 DB 52 LRRHVLINHTLQIQALRLAPAPALPPEPL-----FLGEEFSLSTTIGSLANE 101
 QY 122 L-SHIEGLSQAPPLADEGPPGRSIGAAPSLGALLDLPATGCLLDGEGLE 175
 DB 102 LDTSMYDNLNAPASEGLKPGED--GPGKEAPELDEALDYIMDYVGT 225
 QY 176 DIDTSMYDNLNAPASEGLKPGED--GPGKEAPELDEALDYIMDYVGT 225
 DB 153 DIDTSAVERKVALPPE---PPHSIFCSGSE---WNELDHIMEITIGS 196
 RESULT 12
 Q9UW9 PRELIMINARY; PRT; 196 AA.
 AC Q9UW9:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE RPA-binding trans-activator.
 GN RBR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20440390; PubMed=10982866;
 RA Cho J.M., Song D.J., Alsaouli M.A.;
 RT "Rpt1, a novel transcriptional co-activator, binds the second subunit
 of Replication Protein A.";
 RL Nucleic Acids Res. 28:3478-3485(2000).
 DR EMBL; AF192529; AAF05761.1; -
 SQ SEQUENCE 196 AA; 21798 MW; 33433EF7E8A9EBA7 CRC64;

Query Match 17.6%; Score 213; DB 4; Length 196;
 Best Local Similarity 32.6%; Pred. No. 6.5e-09;

Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

OY 5 GLKRRKEEKEEPLAVDSW-WIDPGHAAVAQAAPPVASSSLFDVSLKHLQSEPD 63
 DB 4 GLKRRHSDLEEE-ERMKSPAGLSYQQA-----LKRISLKVQKSLGPPRAPS 51
 OY 64 LRLHLVYVNTLRRIOAS--MAPAALPPVSPPPAASVADNLASSDAALSASMSLLED 121
 DB 52 LRRHVLHNTLQQLQAALRLAPALPPEPL-----FLGDEDFSLSATIGSILRE 101
 OY 122 L-SHTEGSGAQPQ-----LADGPPGSGSGAASLADLIGATGCLDDGLELFE 175
 DB 102 LDTSDGTEPPONPTVPIGLONEVPP-----QDPVFLKAL--SSRRLGDSGLDDPEFL 152
 OY 176 DITSMYNEMLAPASEGLKPGPED---GPKGEAPELDEALDYIMLVGT 225
 DB 153 DITSAVVEKE---PARAPPEP-PHNLFCAPGSMW-----WNEIDHIMEITIGS 196

RESULT 13

ID 08TB18 PRELIMINARY; PRT; 241 AA.

AC 08TB18; AC 08TB18; AC 08TB18; AC 08TB18;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Skin;
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC025263; AAH25263.1; -
 DR Gene; HGNC:14625; CDCA4.
 KW Hypothetical protein.
 SQ SEQUENCE 241 AA; 26112 MW; CC936EDED4A9A17 CRC64;

Query Match 14.0%; Score 170; DB 4; Length 241;
 Best Local Similarity 28.0%; Pred. No. 1.7e-05;

Matches 74; Conservative 38; Mismatches 90; Indels 62; Gaps 13;

OY 1 MLKSGIKRR--REEEKEPLA---VDSMWLDPGHAAVAQAAPPVASSSLFDVSLKHL 54
 DB 1 MFARGLTKKCYGHEDVGCALAGLKTYSY-----SLQKSLDMSLVKIQ 46
 OY 55 HSLQOSEPDRLHLVYVNTLRRIOASMAPAALPPVSPPPAASVADNLASSDAALSAS 114
 DB 47 LCHMLVEPNLCRSVLIANTVROIEMTQDGTWRTV-APQAERAPLRLVSTELCRAA 105
 OY 115 M-----ASLLEDLSHTEG-----LSQAQPLADG-----PPGRSISGAAPSLG- 153
 DB 106 WQEGCAHPAPGLGD-GHTQGPVSDICPVTSQAAPRHILQSSAMWMDGPRENKGSPFKSLDQ 164
 OY 154 ---ALDLGPATGCLDDGLEDFEDITSYD-NELWAPASEGLKPGPEG----- 201

DB 165 IFETLETNPS--C-----MELFSVDSPYYDITDVTLGMMGARPPCGLEGAPAT 217
 OY 202 PKEEAPELDEALDYIMLVGT 225
 DB 218 PGPSSCKSDLGELDHVEILVET 241

RESULT 14

ID 09BXL8 PRELIMINARY; PRT; 241 AA.

AC 09BXL8; AC 09BXL8; AC 09BXL8; AC 09BXL8;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hematopoietic progenitor protein.
 GN HEPP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21375891; PubMed=11482882;
 RA Abdullah J.M., Jing X., Spassov D.S., Nachman R.G., Juretic R.;
 RT "Cloning and Characterization of Hepp, a Novel Gene Expressed
 Preferentially in Hematopoietic Progenitors and Mature Blood Cells.";
 RL Blood Cells Mol. Dis. 27:667-676(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-B-cell;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF322239; AAK31075.1; -
 DR EMBL; BC011736; AAH11736.1; -
 SQ SEQUENCE 241 AA; 26113 MW; 9693647E77865D7A CRC64;

Query Match 13.6%; Score 165; DB 4; Length 241;
 Best Local Similarity 27.7%; Pred. No. 4.2e-05;

Matches 73; Conservative 39; Mismatches 90; Indels 62; Gaps 13;

OY 1 MLKSGIKRR--REEEKEPLA---VDSMWLDPGHAAVAQAAPPVASSSLFDVSLKHL 54
 DB 1 MFARGLTKKCYGHEDVGCALAGLKTYSY-----SLQKSLDMSLVKIQ 46
 OY 55 HSLQOSEPDRLHLVYVNTLRRIOASMAPAALPPVSPPPAASVADNLASSDAALSAS 114
 DB 47 LCHMLVEPNLCRSVLIANTVROIEMTQDGTWRTV-APQAERAPLRLVSTELCRAA 105
 OY 115 M-----ASLLEDLSHTEG-----LSQAQPLADG-----PPGRSISGAAPSLG- 153
 DB 106 WQEGCAHPAPGLGD-GHTQGPVSDICPVTSQAAPRHILQSSAMWMDGPRENKGSPFKSLDQ 164
 OY 154 ---ALDLGPATGCLDDGLEDFEDITSYD-NELWAPASEGLKPGPEG----- 201
 DB 165 IFETLETNPS--C-----MELFSVDSPYYDITDVTLGMMGARPPCGLEGAPAT 217
 OY 202 PKEEAPELDEALDYIMLVGT 225
 DB 218 PGPSSCKSDLGELDHVEILVET 241
 RESULT 15
 ID 09CWM2 PRELIMINARY; PRT; 237 AA.
 AC 09CWM2; AC 09CWM2; AC 09CWM2; AC 09CWM2;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE ES cells cDNA, RIKEN full-length enriched library, clone:2410018C03
 DE product:hematopoietic progenitor protein, full insert sequence
 DE (Hematopoietic progenitor protein).
 GN Mus musculus (Mouse).
 OS Mus musculus (Mouse).

THIS PAGE BLANK (USPTO)


```
RESULT 2
US-09-083-521-7
; Sequence 7, Application US/09083521
; Patent No. 6048970
; GENERAL INFORMATION:
; APPLICANT: IAL, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,521
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0527 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 853-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1317 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 2459993
; US-09-083-521-7

Query Match      8.8%; Score 107; DB 3; Length 1317;
Best Local Similarity 23.7%; Pred. No. 0.24;
Matches 58; Conservative 20; Mismatches 79; Indels 88; Gaps 10;

OY 78 QASMAPAALPPVSP-----PAAPSVADN-----LLASDA 109
DB 746 EGSAPPOLPLPSVSPSCGASLPSEASAPDILPASPTPAAGSWTVPEPAPTLES5GS 805
OY 110 ALSASMASLIEDLSHTEGLSOAPOPPLADEGPPGRSIC-----GAAPSIGAUDL 157
DB 806 SLGGEAPPS-SEDEDTTEATSEVFTDLSGDPHTKSGIVPALRLSIOKVGCFPDSLSDI 864
OY 158 LGPAT--GC-----LDDGL-----EGLFEDITSMTDNEIMAPAS 191
DB 865 PSSASDGCCEVLSPSAGPPGCGPRAVDGVDTEYTESPEFVLKHAHSSPEAFGEFAS 924
OY 192 EGLMPGPD-----GPKKEAPELDEA---ELDYIMDVLYG-----TQALR 230
DB 925 EGESBGPPLLSVLSIGLSKSPYRDSAYFSDLDNESEPTFGPEKHSGIODSOKEQDLRS 984
OY 231 PPGPG 235
DB 985 PPSPG 989
```

```
RESULT 3
US-09-252-991A-27790
; Sequence 27790, Application US/09252991A
; Patent No. 6551795
```

```
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27790
; TYPE: PRT
; LENGTH: 760
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-27790

Query Match      8.8%; Score 106.5; DB 4; Length 760;
Best Local Similarity 25.7%; Pred. No. 0.13;
Matches 59; Conservative 29; Mismatches 101; Indels 41; Gaps 10;

OY 2 LSKGLKRRKEE--EKEPLAVDSMWLDPRG-AAVAQAPPAVASSLSFLSYKLHHSIQ 58
DB 126 LLAALRLAEPEGAPPAEPVQAPPAVPAPEPAEPVQAPASASSDITDDEFQLDALQ 185
OY 59 QSEPDRLVLVVTLRRIQASMAPAALPPVSPPAAPSVADNLLASDAALSASMASL 118
DB 186 GDE-----APASAVAPAPAPAGDESD-----APEAL 214
OY 119 LEDSHTEGLSOAPOPPLADEGP-PGRSTGGAAPSLGALDLGPRATGCLLD--GLEGLFE 175
DB 215 LDQL-HGKGFVPPAVSAEPQVAPAEVAPAAAGD-DISDDEFALLDELHG-KGKFG 271
OY 176 DIDTSMYDNEIMAPASEGLKPGPE-DGPKKEAPELDEAEIDYLDVLYG 224
DB 272 DYPEAGTPAPAPAAAAPAAAEQKAPAAAGGEISDDEFESLDELHG 321

RESULT 4
US-08-795-868-14
; Sequence 14, Application US/08795868
; Patent No. 5845773
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC
; TITLE OF INVENTION: AND STRIPED-SPECIFIC MUSCLE CELL ISOFORMS AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,868
; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/494,577
; FILING DATE: 22-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/032001
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: 617-542-5070
 TELEFAX: 617-542-8906
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 661 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 US-08-795-868-14

Query Match 8.6%; Score 104; DB 2; Length 661;
 Best Local Similarity 21.9%; Pred. No. 0.18;

Matches 73; Conservative 28; Mismatches 104; Indels 128; Gaps 15;

QY 7 KRKEEKEKEPLAVSW-----WIDPGHA-AVAQAPPAVAS-SSL 45
 DB 109 RRRSLERSDPPAPLPWPLRKARSLERQPKSERGAPMGTGASQDELPAQSVARRRL 168
 QY 46 FDLVLKHLHSLQOSEP---DLR-----HLYLVNTLR----- 75
 DB 169 FQOKASLDERTRQSRPASDELRFQDELGRTRRSTRRELYVSHESLRTATIQRAPSPRE 228
 QY 76 -----RQASMAPAALPPVPSPPAAPSVADNLASSDA----- 110
 DB 229 PGEPLFSRPTPKTSRAVSPAAPSPSSAEKPEDDEGRPRSRGPA RTEPGEPPQE 288
 QY 111 -----LSASMASLLEDLSHIEGLSQAPOLAD-----EGPPGRSIGGAAPSIGALD 156
 DB 289 VRRDOFPLTRSRALQ-----ECRSVPVPPAADPEARTKAPGRK---REPPAQA VR 338
 QY 157 LGGPATGCLLDGLEGLEFEDIDTSMYDNELMAPASGLKRP-EDGP----- 202
 DB 339 FLPMAT-----PGLBG--AAVPTLEKNRAGPEAEKRLRGRPEDEGPMGPMWRGARSOG 391
 QY 203 ---GKEAPELDEALDYIMDVLTQALERP 231
 DB 392 KGRRAPTSPELESSDSY---VSAGEEPLERP 421

RESULT 5

US-09-303-069-14
 ; Sequence 14, Application US/09303069A
 ; Patent No. 6350592
 ; GENERAL INFORMATION:
 ; APPLICANT: Lee, Mu-Ba
 ; APPLICANT: Hsieh, Chung-Ming
 ; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
 ; TITLE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
 ; FILE REFERENCE: 05433/039001
 ; CURRENT APPLICATION NUMBER: US/09/303,069A
 ; CURRENT FILING DATE: 1999-04-30
 ; EARLIER APPLICATION NUMBER: US 09/134,250
 ; EARLIER FILING DATE: 1998-08-14
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 14
 ; LENGTH: 661
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-303-069-14

Query Match 8.6%; Score 104; DB 4; Length 661;
 Best Local Similarity 21.9%; Pred. No. 0.18;

Matches 73; Conservative 28; Mismatches 104; Indels 128; Gaps 15;

QY 7 KRKEEKEKEPLAVDSW-----WIDPGHA-AVAQAPPAVAS-SSL 45
 DB 109 RRRSLERSDPPAPLPWPLRKARSLERQPKSERGAPMGTGASQDELPAQSVARRRL 168
 QY 46 FDLVLKHLHSLQOSEP---DLR-----HLYLVNTLR----- 75

DB 169 FQOKASLDERTRQSRPASDELRFQDELGRTRRSTRRELYVSHESLRTATIQRAPSPRE 228
 QY 76 -----RQASMAPAALPPVPSPPAAPSVADNLASSDA----- 110
 DB 229 PGEPLFSRPTPKTSRAVSPAAPSPSSAEKPEDDEGRPRSRGPA RTEPGEPPQE 288
 QY 111 -----LSASMASLLEDLSHIEGLSQAPOLAD-----EGPPGRSIGGAAPSIGALD 156
 DB 289 VRRDOFPLTRSRALQ-----ECRSVPVPPAADPEARTKAPGRK---REPPAQA VR 338
 QY 157 LGGPATGCLLDGLEGLEFEDIDTSMYDNELMAPASGLKRP-EDGP----- 202
 DB 339 FLPMAT-----PGLBG--AAVPTLEKNRAGPEAEKRLRGRPEDEGPMGPMWRGARSOG 391
 QY 203 ---GKEAPELDEALDYIMDVLTQALERP 231
 DB 392 KGRRAPTSPELESSDSY---VSAGEEPLERP 421

RESULT 6

US-09-134-250-14
 ; Sequence 14, Application US/09134250B
 ; Patent No. 639753
 ; GENERAL INFORMATION:
 ; APPLICANT: Lee, Mu-Ba
 ; APPLICANT: Hsieh, Chung-Ming
 ; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
 ; TITLE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
 ; FILE REFERENCE: 05433/038001
 ; CURRENT APPLICATION NUMBER: US/09/134,250B
 ; CURRENT FILING DATE: 1998-08-14
 ; EARLIER APPLICATION NUMBER: 1997-02-06
 ; EARLIER FILING DATE: 1997-02-06
 ; EARLIER APPLICATION NUMBER: US 08/494,577
 ; EARLIER FILING DATE: 1995-06-22
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 14
 ; LENGTH: 661
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-134-250-14

Query Match 8.6%; Score 104; DB 4; Length 661;
 Best Local Similarity 21.9%; Pred. No. 0.18;

Matches 73; Conservative 28; Mismatches 104; Indels 128; Gaps 15;

QY 7 KRKEEKEKEPLAVDSW-----WIDPGHA-AVAQAPPAVAS-SSL 45
 DB 109 RRRSLERSDPPAPLPWPLRKARSLERQPKSERGAPMGTGASQDELPAQSVARRRL 168
 QY 46 FDLVLKHLHSLQOSEP---DLR-----HLYLVNTLR----- 75
 DB 169 FQOKASLDERTRQSRPASDELRFQDELGRTRRSTRRELYVSHESLRTATIQRAPSPRE 228
 QY 76 -----RQASMAPAALPPVPSPPAAPSVADNLASSDA----- 110
 DB 229 PGEPLFSRPTPKTSRAVSPAAPSPSSAEKPEDDEGRPRSRGPA RTEPGEPPQE 288
 QY 111 -----LSASMASLLEDLSHIEGLSQAPOLAD-----EGPPGRSIGGAAPSIGALD 156
 DB 289 VRRDOFPLTRSRALQ-----ECRSVPVPPAADPEARTKAPGRK---REPPAQA VR 338
 QY 157 LGGPATGCLLDGLEGLEFEDIDTSMYDNELMAPASGLKRP-EDGP----- 202
 DB 339 FLPMAT-----PGLBG--AAVPTLEKNRAGPEAEKRLRGRPEDEGPMGPMWRGARSOG 391
 QY 203 ---GKEAPELDEALDYIMDVLTQALERP 231
 DB 392 KGRRAPTSPELESSDSY---VSAGEEPLERP 421

RESULT 7


```
RESULT 10
US-09-568-480-5
; Sequence 5, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-5
```

```
Query Match      8.6%; Score 103.5; DB 4; Length 7257;
Best Local Similarity 24.2%; Pred. No. 5.1;
Matches 66; Conservative 24; Mismatches 98; Indels 85; Gaps 10;
```

```
QY 31 AAVAQAPPAVASSSLEFDL--SVYKLHNSLQOSEPDRLHL--VLVYNTL----- 74
DB 3365 AAKANYAHLEAASGLASLTKAVLALRHEQIPAPPELGEINPMLPWMTLVVAAPRAKAVPWG 3424
QY 75 -----RR-----IQASMAPAALPPVSPAPASVADNLIASSDAALSASMAS 117
DB 3425 RGAAPRAAGVSAFGLSGTIVHVVLEAPVEEPAPAPARVELVYLSAKSAALDAAAR 3484
QY 118 LLEDUS-HIE-----GLSQAPDPL-----ADGGPPGRSIG 146
DB 3485 LSAHLSAHPLELSIGDAVFSLATTRSPMEHRLAATTSREALRGALDAAAOQKTPGAVRG 3544
QY 147 GAAPSIGALDLGPATGCLLDGLEGLEF-----DIDTSMYDNEIMAPASEGLKPG 197
DB 3545 KAVSSRGKLAFLFTGGGAQMPGKRGLEYETWPAFRFAFRVCYALFDRIDQPLREVMAA- 3603
QY 198 PEDGPKKEAPPELDEA-----ELDYLMQVYL 222
DB 3604 ---APGLAQAAARLDQTAAYAPALFALEVALAAL 3633
```

```
RESULT 11
US-09-568-486-5
; Sequence 5, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
```

```
; ORGANISM: Sorangium cellulosum
US-09-568-486-5
```

```
Query Match      8.6%; Score 103.5; DB 4; Length 7257;
Best Local Similarity 24.2%; Pred. No. 5.1;
Matches 66; Conservative 24; Mismatches 98; Indels 85; Gaps 10;
```

```
QY 31 AAVAQAPPAVASSSLEFDL--SVYKLHNSLQOSEPDRLHL--VLVYNTL----- 74
DB 3365 AAKANYAHLEAASGLASLTKAVLALRHEQIPAPPELGEINPMLPWMTLVVAAPRAKAVPWG 3424
QY 75 -----RR-----IQASMAPAALPPVSPAPASVADNLIASSDAALSASMAS 117
DB 3425 RGAAPRAAGVSAFGLSGTIVHVVLEAPVEEPAPAPARVELVYLSAKSAALDAAAR 3484
QY 118 LLEDUS-HIE-----GLSQAPDPL-----ADGGPPGRSIG 146
DB 3485 LSAHLSAHPLELSIGDAVFSLATTRSPMEHRLAATTSREALRGALDAAAOQKTPGAVRG 3544
QY 147 GAAPSIGALDLGPATGCLLDGLEGLEF-----DIDTSMYDNEIMAPASEGLKPG 197
DB 3545 KAVSSRGKLAFLFTGGGAQMPGKRGLEYETWPAFRFAFRVCYALFDRIDQPLREVMAA- 3603
QY 198 PEDGPKKEAPPELDEA-----ELDYLMQVYL 222
DB 3604 ---APGLAQAAARLDQTAAYAPALFALEVALAAL 3633
```

```
RESULT 12
US-09-568-472-5
; Sequence 5, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-5
```

```
Query Match      8.6%; Score 103.5; DB 4; Length 7257;
Best Local Similarity 24.2%; Pred. No. 5.1;
Matches 66; Conservative 24; Mismatches 98; Indels 85; Gaps 10;
```

```
QY 31 AAVAQAPPAVASSSLEFDL--SVYKLHNSLQOSEPDRLHL--VLVYNTL----- 74
DB 3365 AAKANYAHLEAASGLASLTKAVLALRHEQIPAPPELGEINPMLPWMTLVVAAPRAKAVPWG 3424
QY 75 -----RR-----IQASMAPAALPPVSPAPASVADNLIASSDAALSASMAS 117
DB 3425 RGAAPRAAGVSAFGLSGTIVHVVLEAPVEEPAPAPARVELVYLSAKSAALDAAAR 3484
QY 118 LLEDUS-HIE-----GLSQAPDPL-----ADGGPPGRSIG 146
DB 3485 LSAHLSAHPLELSIGDAVFSLATTRSPMEHRLAATTSREALRGALDAAAOQKTPGAVRG 3544
QY 147 GAAPSIGALDLGPATGCLLDGLEGLEF-----DIDTSMYDNEIMAPASEGLKPG 197
DB 3545 KAVSSRGKLAFLFTGGGAQMPGKRGLEYETWPAFRFAFRVCYALFDRIDQPLREVMAA- 3603
```

```
QY      198 PEDGGKEAPELDEA-----ELDYIMDVL 222
          || : | || :           | : | : |
Db      3604 --APGLAQAAARDQTAYAQPALFALEYALAAAL 3633
```

RESULT 13
US-09-567

```

Sequence 5, Application US/09567899
Patent No. 6383787
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zickle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30562A
CURRENT APPLICATION NUMBER: US/09/567, 899
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 7257
TYPE: PRF
ORGANISM: Sorangium cellulosum
US-09-367-899-5

```

| | | | | |
|-----------------------|------------------|----------------|------------|--------------|
| Query Match | 8.6% | Score 103.5; | DB 4; | Length 7257; |
| Best Local Similarity | 24.2% | Pred. No. 5.1; | | |
| Matches 66; | Conservative 24; | Mismatches 98; | Indels 85; | Gaps 10 |

| | | | |
|----|------|---|------|
| QY | 31 | AAVQAAPPAAVSSSLFDL--SVLKHHLSLOOSEPLRLH--VLIVNTL----- | 74 |
| Dd | 3365 | AARKNVAHLEAASGLSILKRAVLNLRHQIIPAQPELGSLNHLPNNTLPPAVPRKAAYPMWG | 3422 |
| QY | 75 | ----RR-----IQASMAPPAALPVPSPAPASVADNLIASDPAALSAMAS | 117 |
| Dd | 3425 | RGARPRRAGVSASFGLSTGNVHVLEEAPVEPAPAAPARPVELVYLISAKSAALDLDAAR | 3489 |
| QY | 118 | LLEDLS-HIE-----GLSQAQPL-----ADEGPGRSIG | 146 |
| Dd | 3485 | LSAHLASHPELISLDVAFSLATITRSPMHRLLAITTSREALRGALDAANAOOTPGGAIRG | 3544 |
| QY | 147 | GAAPSLGALDLIGPATGCLDDDGLEGLEFE-----DIDTSMYNELMAIPASBGLKPG | 197 |
| Dd | 3545 | KAVSSRKLAFLFTGGQAGMPGMGRGLGYETPPAFREAFDRCVALTFRDIIDGPLREVMMA- | 3603 |
| QY | 198 | PEDFGKEAEARELDEA-----ELDTIMDYLV | 222 |
| Dd | 3604 | ---APGIAGAARLDQTAYAAPALFALEATAUAL | 3633 |

RESULT 14
US-08-857

Sequence 57, Application US/08857076C
 Patent No. 6225120
 GENERAL INFORMATION:
 APPLICANT: Ruvkun, Gary
 APPLICANT: Kimura, Koutaro
 APPLICANT: Patterson, Garth
 APPLICANT: Ogg, Scott
 APPLICANT: Paradis, Suzanne
 APPLICANT: Tiesenbaum, Heidi
 APPLICANT: Morris, Jason
 APPLICANT: Kowek, Allison
 TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
 TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
 FILE REFERENCE: 00786/351001
 CURRENT APPLICATION NUMBER: US/08/857,076C
 CURRENT FILING DATE: 1997-05-15

```

: NUMBER OF SEQ ID NOS: 114
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 57
: LENGTH: 655
: TYPE: PRT
: ORGANISM: Homo sapiens
: OS-08-857-076-57

```

| | | | | |
|--------------------------|--------|-----------------|-----------|-------------|
| Query Match | 8.3% | Score 100; | DB 3; | Length 655; |
| Best Local Similarity | 36.5%; | Pred. No. 0.42; | | |
| Matches 35; Conservative | 9; | Mismatches 44; | Indels 8; | Gaps 3; |

QY 74 LRRIQASMAPAAALPVPSPPAA--PVSADNKLISDPAALSAMSLLEDLSHIEGLSQQA 131
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 27 LPREFEQSNGSATSSPAPSGSAANPDAAAGLPASAAVAADFMSNTSLLEESDPFQA 86

QY 132 PQRLADEGPRGRSTCGAAPSLGAL--DLGGPATGCL 165
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 87 PGSYA-----AAVAAAAAAAAATGCLCGDPGGEGRGL 118

RESULT 15
US-09-345

```

: Sequence 40, Application US/09345473E
: Patent No. 6558903
:
: GENERAL INFORMATION:
:
: APPLICANT: Hodge, Martin
:
: TITLE OF INVENTION: No. 6558903el Kinases and Uses Thereof
:
: FILE REFERENCE: 35800/183781
:
: CURRENT APPLICATION NUMBER: US/09/345,473E
:
: CURRENT FILING DATE: 1999-06-30
:
: NUMBER OF SEQ ID NOS: 62
:
: SOFTWARE: FASTSEQ for Windows Version 4.0
:
: SEQ ID NO 40
:
: LENGTH: 1601
:
: TYPE: PRF
:
: ORGANISM: C. elegans
:
: US-09-345-473E-40

```

| | | | | |
|-----------------------|------------------|----------------|------------|--------------|
| Query Match | 8.3%; | Score 100; | DB 4; | Length 1601; |
| Best Local Similarity | 25.1%; | Pred. NO. 1.4; | | |
| Matches 64; | Conservative 29; | Mismatches 92; | Indels 70; | Gaps 14; |

[illegible]

Search completed: August 19, 2003, 09:37:02
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:35:04 ; Search time 57 Seconds
(without alignments)
542.399 Million cell updates/sec

Title: US-09-701-675A-3
 Perfect score: 1210
 Sequence: 1 MLKGLKRRKEEEKPLA.....YLMDLVGTQALERPPGGR 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 492763 seqs, 131003257 residues
Total number of hits satisfying chosen parameters: 492763

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
```

```

1:  published_applications_AA.*
2:  /cgn2_6/p/odataa/2/pubppaa/US07_PUBCOMB.pdep.*
3:  /cgn2_6/p/odataa/2/pubppaa/PCT_NEW_PUB.pdep.*
4:  /cgn2_6/p/odataa/2/pubppaa/US06_NEW_PUB.pdep.*
5:  /cgn2_6/p/odataa/2/pubppaa/US06_PUBCOMB.pdep.*
6:  /cgn2_6/p/odataa/2/pubppaa/US07_NEW_PUB.pdep.*
7:  /cgn2_6/p/odataa/2/pubppaa/US08_NEW_PUB.pdep.*
8:  /cgn2_6/p/odataa/2/pubppaa/US08_PUBCOMB.pdep.*
9:  /cgn2_6/p/odataa/2/pubppaa/US09A_PUBCOMB.pdep.*
10: /cgn2_6/p/odataa/2/pubppaa/US09B_PUBCOMB.pdep.*
11: /cgn2_6/p/odataa/2/pubppaa/US09C_PUBCOMB.pdep.*
12: /cgn2_6/p/odataa/2/pubppaa/US09_NEW_PUB.pdep.*
13: /cgn2_6/p/odataa/2/pubppaa/US10A_PUBCOMB.pdep.*
14: /cgn2_6/p/odataa/2/pubppaa/US10B_PUBCOMB.pdep.*
15: /cgn2_6/p/odataa/2/pubppaa/US10C_PUBCOMB.pdep.*
16: /cgn2_6/p/odataa/2/pubppaa/US10_NEW_PUB.pdep.*
17: /cgn2_6/p/odataa/2/pubppaa/US06_NEW_PUB.pdep.*
18: /cgn2_6/p/odataa/2/pubppaa/US06_PUBCOMB.pdep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | | | | | Description |
|--------|-------|-------|--------|----|---------------------|---------------------|
| No. | Score | Match | Length | DB | ID | |
| 1 | 1206 | 99.7 | 236 | 10 | US-09-731-872-412 | Sequence 412, App |
| 2 | 168 | 13.9 | 254 | 9 | US-09-923-302-596 | Sequence 596, App |
| 3 | 168 | 13.9 | 254 | 15 | US-10-106-698-4456 | Sequence 4456, App |
| 4 | 165 | 13.6 | 241 | 14 | US-10-076-069-4 | Sequence 4, App |
| 5 | 163.5 | 13.5 | 157 | 15 | US-10-106-698-5516 | Sequence 5516, App |
| 6 | 152 | 12.6 | 237 | 14 | US-10-076-069-2 | Sequence 2, App |
| 7 | 115 | 9.5 | 255 | 10 | US-09-764-846-139 | Sequence 139, App |
| 8 | 115 | 9.5 | 255 | 15 | US-10-091-483-139 | Sequence 139, App |
| 9 | 113 | 9.3 | 983 | 9 | US-09-912-917-2 | Sequence 2, App |
| 10 | 112 | 9.3 | 715 | 15 | US-10-157-031-115 | Sequence 115, App |
| 11 | 111 | 9.1 | 1259 | 15 | US-10-260-715-8 | Sequence 8, App |
| 12 | 108.5 | 9.0 | 803 | 15 | US-10-156-761-12958 | Sequence 12958, App |
| 13 | 107 | 8.8 | 1317 | 10 | US-09-963-896-7 | Sequence 7, App |
| 14 | 104 | 8.6 | 113 | 14 | US-10-076-069-6 | Sequence 6, App |
| 15 | 104 | 8.6 | 459 | 9 | US-09-789-561-97 | Sequence 97, App |

| | | | | | | |
|----|-------|-----|------|----|---------------------|--------------------|
| 16 | 104 | 8.6 | 661 | 14 | US-10-160-865-14 | Sequence 14, App1 |
| 17 | 103.5 | 8.6 | 1567 | 9 | US-09-835-232-2 | Sequence 2, App1 |
| 18 | 103.5 | 8.6 | 7257 | 14 | US-10-014-717-5 | Sequence 5, App1 |
| 19 | 102.5 | 8.5 | 674 | 15 | US-10-156-761-9618 | Sequence 9618, App |
| 20 | 102 | 8.4 | 1516 | 15 | US-10-060-036-166 | Sequence 166, App |
| 21 | 101 | 8.3 | 433 | 10 | US-09-801-368-38 | Sequence 38, App1 |
| 22 | 101 | 8.3 | 1497 | 15 | US-10-157-031-128 | Sequence 128, App |
| 23 | 100 | 8.3 | 655 | 9 | US-03-205-658-57 | Sequence 57, App1 |
| 24 | 100 | 8.3 | 655 | 9 | US-09-844-353A-47 | Sequence 57, App1 |
| 25 | 100 | 8.3 | 1601 | 10 | US-09-862-027-40 | Sequence 40, App1 |
| 26 | 99.5 | 8.2 | 117 | 14 | US-10-076-069-7 | Sequence 7, App1 |
| 27 | 99.5 | 8.2 | 559 | 15 | US-10-156-761-11383 | Sequence 11383, A |
| 28 | 99.5 | 8.2 | 682 | 15 | US-10-156-761-12088 | Sequence 12088, A |
| 29 | 99.5 | 8.2 | 1232 | 12 | US-10-017-161-2396 | Sequence 2395, App |
| 30 | 99.5 | 8.2 | 4471 | 15 | US-10-205-032-10 | Sequence 10, App1 |
| 31 | 98.5 | 8.1 | 300 | 15 | US-10-156-761-8940 | Sequence 8940, App |
| 32 | 97.5 | 8.1 | 830 | 15 | US-10-211-962-32 | Sequence 32, App1 |
| 33 | 97 | 8.0 | 903 | 15 | US-10-156-761-11093 | Sequence 11093, A |
| 34 | 96.5 | 8.0 | 1244 | 15 | US-10-156-761-7652 | Sequence 7652, App |
| 35 | 96 | 7.9 | 641 | 15 | US-10-156-761-7944 | Sequence 7944, App |
| 36 | 96 | 7.9 | 1298 | 10 | US-09-825-2888-2 | Sequence 2, App1 |
| 37 | 95 | 7.9 | 982 | 9 | US-09-841-132-176 | Sequence 176, App |
| 38 | 95 | 7.9 | 1006 | 9 | US-09-841-132-190 | Sequence 190, App |
| 39 | 95 | 7.9 | 1265 | 10 | US-09-862-032-80 | Sequence 80, App1 |
| 40 | 94.5 | 7.8 | 830 | 15 | US-10-211-962-31 | Sequence 31, App1 |
| 41 | 94 | 7.8 | 457 | 15 | US-10-078-770-1194 | Sequence 194, App1 |
| 42 | 93.5 | 7.7 | 582 | 15 | US-10-156-761-11466 | Sequence 11466, A |
| 43 | 93.5 | 7.7 | 1203 | 14 | US-10-027-923-5 | Sequence 5, App1 |
| 44 | 93.5 | 7.7 | 2971 | 15 | US-10-146-473-50 | Sequence 50, App1 |
| 45 | 93 | 7.7 | 267 | 15 | US-10-156-761-11403 | Sequence 11403, A |

ALIGNMENTS

```

RESULT 1
US-09-731-872-412
: Sequence 412, Application US/09731872
: Patent No. US20020102604A1
:
: GENERAL INFORMATION:
: APPLICANT: Dumas Milne Edwards, Jean Baptiste
: APPLICANT: Bougueleret, Lydie
: APPLICANT: Jobert, Severin
: TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
: FILE REFERENCE: 78 US3 REG
: CURRENT APPLICATION NUMBER: US/09/731,872
: CURRENT FILING DATE: 2000-12-07
: PRIOR APPLICATION NUMBER: US 60/169,639
: PRIOR FILING DATE: 1999-12-08
: PRIOR APPLICATION NUMBER: US 60/187,470
: PRIOR FILING DATE: 2000-03-06
: NUMBER OF SEQ ID NOS: 482
: SOFTWARE: Patent.pm
: SEQ ID NO 412
: LENGTH: 236
:
: TYPE: prt
: ORGANISM: Homo sapiens
: US-09-731-872-412

```

| | | | | |
|-----------------------|-----------------|---|-----------|-------------|
| Query Match | 99.7% | Score 1206: | DB 10: | Length 236: |
| Best Local Similarity | 99.6% | Pred. No 2.4e-93: | | |
| Matches 235: | Conservative 0: | Mismatches 1: | Indels 0: | Gaps 0: |
| QY | 1 | MLSGKLRKREEEKEEPLAVDSNWLDPGHAAVAQAPPAVASSSLFDLSVTLKHLHSLOOS | 60 | |
| | | | | |
| Db | 1 | MLSGKLRKREEEKEEPLAVDSNWLDPGHAAVAQAPPAVASSSLFDLSVTLKHLHSLOOS | 60 | |
| QY | 61 | EPDLRHLVLYVNTLRRIQASMAAPPAALPVPVSPAPPAEVSANLTLASSDAALSASMASTLE | 120 | |
| | | | | |
| Db | 61 | EPDLRHLVLYVNTLRRIQASMAAPPAALPVPVSPAPPAEVSANLTLASSDAALSASMASTLE | 120 | |
| QY | 121 | DLSHIEGLISGAOPDLADEGPPGRSIRGGAAPSLGALDILGPATGCLLDLGLTEGLEFEEDIDTS | 180 | |
| | | | | |

Db 121 DISHTEGSQLAPOLADEGPPGRSIGGAAPSLGALDLPATGCLLDGLEGLEDITDTS 180
 QY 181 MIDNELMAPASEGLKPGEDGPGKEAPDELDAEILDYLDVLTGQALERPGR 236
 Db 181 MIDNELMAPASEGLKPGEDGPGKEAPDELDAEILDYLDVLTGQALERPGR 236

RESULT 2 US-09-925-302-596

; Sequence 596, Application US/09925302
 ; Patent No. US20020044941A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA104
 ; CURRENT APPLICATION NUMBER: US/09/925,302
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05918
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 896
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 596
 ; LENGTH: 254
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (105)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-925-302-596

Query Match 13.9%; Score 168; DB 9; Length 254;
 Best Local Similarity 27.7%; Pred. No. 2,3e-06;
 Matches 73; Conservative 39; Mismatches 90; Indels 62; Gaps 13;

QY 1 MISKGLKRR--REEEKEEPLA-----VDSMWLDPGHAAVQAQAPPAVSSSLFDLSVLKLIH 54
 Db 14 MFARGLKRCVGHEDVDEGALAGLKTVSSY-----SLQRSLDMSLVKIQ 59
 QY 55 HSLQSEPDRLHVLVVTLRRIQASMAPAALPPVSPAPASVADNLTASSDAALSAS 114
 Db 60 LCHMLVEPNLCRSVLIANTVQIQEEMTQDGTWRTV-APQAARAPADRLVSTELICRAA 118
 QY 115 M-----ASLEDLSHTEG-----LSQAQPLADEG-----PGRSIGGAAPSLG- 153
 Db 119 WQEGAHAPAGLGD-GHTQGPVSDLCPYTSAQAPRHLSASMEMDGPRENKGFHKSLDQ 177
 QY 154 ---ALDLGPARTCGLLDGLEGLEDITDTSMTD-NELMAPASEGLKPGEDG----- 201
 Db 178 IFETLETNPSS--C-----MELEFSDVSPYDIDTVLTGMWGARGAPGPCGLEGLAPAT 230
 QY 202 PGKEAPELDEAEILDYLDVLTGVT 225
 Db 231 PGSSCKSDLGELDHVEILVET 254

RESULT 3

US-10-106-698-4456
 ; Sequence 4456, Application US/10106698
 ; Publication No. US20030109690A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Colon- and Colon Cancer Associated Polynucleotides and Polypeptide
 ; FILE REFERENCE: PA005P1
 ; CURRENT APPLICATION NUMBER: US/10/106,698
 ; PRIOR FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524
 ; PRIOR FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: US 60/157,137
 ; PRIOR FILING DATE: 1999-09-29
 ; PRIOR APPLICATION NUMBER: US 60/163,280

; PRIOR FILING DATE: 1999-11-03
 ; NUMBER OF SEQ ID NOS: 8564
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 4456
 ; LENGTH: 254
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (105)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-10-106-698-4456

Query Match 13.9%; Score 168; DB 15; Length 254;
 Best Local Similarity 27.7%; Pred. No. 2,3e-06;
 Matches 73; Conservative 39; Mismatches 90; Indels 62; Gaps 13;

QY 1 MISKGLKRR--REEEKEEPLA-----VDSMWLDPGHAAVQAQAPPAVSSSLFDLSVLKLIH 54
 Db 14 MFARGLKRCVGHEDVDEGALAGLKTVSSY-----SLQRSLDMSLVKIQ 59
 QY 55 HSLQSEPDRLHVLVVTLRRIQASMAPAALPPVSPAPASVADNLTASSDAALSAS 114
 Db 60 LCHMLVEPNLCRSVLIANTVQIQEEMTQDGTWRTV-APQAARAPADRLVSTELICRAA 118
 QY 115 M-----ASLEDLSHTEG-----LSQAQPLADEG-----PGRSIGGAAPSLG- 153
 Db 119 WQEGAHAPAGLGD-GHTQGPVSDLCPYTSAQAPRHLSASMEMDGPRENKGFHKSLDQ 177
 QY 154 ---ALDLGPARTCGLLDGLEGLEDITDTSMTD-NELMAPASEGLKPGEDG----- 201
 Db 178 IFETLETNPSS--C-----MELEFSDVSPYDIDTVLTGMWGARGAPGPCGLEGLAPAT 230
 QY 202 PGKEAPELDEAEILDYLDVLTGVT 225
 Db 231 PGSSCKSDLGELDHVEILVET 254

RESULT 4

US-10-076-069-4
 ; Sequence 4, Application US/10076069
 ; Publication No. US20020177214A1
 ; GENERAL INFORMATION:
 ; APPLICANT: JURECIC, ROLAND
 ; TITLE OF INVENTION: HPP, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEUTRAL DEV
 ; FILE REFERENCE: 39532-176599
 ; CURRENT APPLICATION NUMBER: US/10/076,069
 ; PRIOR FILING DATE: 2002-02-15
 ; PRIOR APPLICATION NUMBER: US 60/268,923
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 241
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-076-069-4

Query Match 13.6%; Score 165; DB 14; Length 241;
 Best Local Similarity 27.7%; Pred. No. 3,9e-06;
 Matches 73; Conservative 39; Mismatches 90; Indels 62; Gaps 13;

QY 1 MISKGLKRR--REEEKEEPLA-----VDSMWLDPGHAAVQAQAPPAVSSSLFDLSVLKLIH 54
 Db 1 MFARGLKRCVGHEDVDEGALAGLKTVSSY-----SLQRSLDMSLVKIQ 46
 QY 55 HSLQSEPDRLHVLVVTLRRIQASMAPAALPPVSPAPASVADNLTASSDAALSAS 114
 Db 47 LCHMLVEPNLCRSVLIANTVQIQEEMTQDGTWRTV-APQAARAPADRLVSTELICRAA 105
 QY 115 M-----ASLEDLSHTEG-----LSQAQPLADEG-----PGRSIGGAAPSLG- 153

Db 106 WGOEGHAPAPGLCD-GHTGCVSDICPVMSQAQPHRLQSSAWEMDGPRENRSFHKSLDQ 164
QY 154 ---ALDLGPATCLDLDGLEGLEDIDTSMYD-NELMAPASEGLKPGEDG----- 201
Db 165 IFETLETKNPS--C-----MEELFSDVDSPIYDLDVVLGMGAGARPGCEGLEAPAT 217
QY 202 PGKEAPELDEALDYLMQVYGT 225
Db 218 PGSSSCSKSDGLGELDHVEILVET 241

RESULT 5

US-10-106-698-5516
; Sequence 5516, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106, 698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5516
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (55)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (132)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5516

Query Match 13.5%; Score 163.5; DB 15; Length 157;
Best Local Similarity 32.6%; Pred. No. 3e-06;
Matches 58; Conservative 27; Mismatches 48; Indels 45; Gaps 10;

QY 62 PDLRHLVYVNTLRIOAS--MAPAAALPPVSPAPASVADNLLASSDAASASASL 119
Db 11 PSIRRVVLIHNTLQQLQALRLAPALPPEPL-----FLGEEDPSLSAKIGSL 60
QY 120 EDL-SHIEGLSQAPQ---LADGPPGRSIGGAAPSIGALDLDGPATGCLDGLGL 173
Db 61 RELDTSMDSTEPQNVPTPLGONEVPP-----QPDVFLEAL--SSRYIGDSGLD 111
QY 174 FEDITSMDNLMAPASGLKPGED-----GCKEKAPELDEALDYLMQVYGT 225
Db 112 FLDIDTSAVEKE--PA---RAPEPXMHLFCAPGSWE---WNELDHIMEIILGS 157

RESULT 6

US-10-076-069-2
; Sequence 2, Application US/10076069
; Publication No. US2002017721A1
; GENERAL INFORMATION:
; APPLICANT: JURECIC, ROLAND
; APPLICANT: NACHTMAN, RONALD
; TITLE OF INVENTION: HEP, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEVELO
; FILE REFERENCE: 39532-176399
; CURRENT APPLICATION NUMBER: US/10/076, 069
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/268, 923
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-076-069-2

Query Match 12.6%; Score 152; DB 14; Length 237;
Best Local Similarity 26.8%; Pred. No. 4.7e-05;
Matches 70; Conservative 35; Mismatches 96; Indels 60; Gaps 10;

QY 1 MSLGKLRKEBEKEPLAVDSWILDPGHAAYQAAPPAVASSFLDLSVLKLSHQSS 60
Db 1 MFARGIKRKYGDQEE---GVGFGFVPSYSIQRQ-----SLIDMSLVKIQLCMLV 48
QY 61 EPDLRHLVYVNTLRIOASMAP-----AAALPPVSPAPASVADNLLASS-DALSS 114
Db 49 EPMLCSTVILANTVRIQEMSGODGVHGMAPONDRALVERLVSTELLCTVVGAEER 108
QY 115 MASLLEDLSHISGLSQAPQPLADEGPPGRSIGAAP-----SIGALDLDGPATGCL-- 165
Db 109 PAPELEDAPLQNSVSELP-----IYGSAPGCRNPPSSLSMEMDSPENKGSFQK 156
QY 166 -LD-----DGLGLFEDIDTSMYD-NELMAPASEGLKPGEDG-----PGK 204
Db 157 SLDQIFETLENKNSSSVEELFSDVSSYYDLDVVLGMGSKRSICNGLGPAATPPP 216
QY 205 EEAPELDEALDYLMQVYGT 225
Db 217 SSTCKSDLAELDHVEILVET 237

RESULT 7

US-09-764-846-139
; Sequence 139, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212
; CURRENT APPLICATION NUMBER: US/09/764, 846
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 139
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (19)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-846-139

Query Match 9.5%; Score 115; DB 10; Length 255;
Best Local Similarity 26.6%; Pred. No. 0.065;
Matches 68; Conservative 27; Mismatches 83; Indels 79; Gaps 14;

QY 28 PGHAAYQAAP-----AVASSFLDLSVLKLSHQSSPDLDHLVYVNT 73
Db 21 PGGAADASLPGDAGTRGGGGLYSKAPNEMWRCY-----RRKLSLFLPLPINA 72
QY 74 LRR--IQASMAPAALPPVSPAPASVADNLLASSDAA-ISASNASLLEDLSHIEGLSQ 130

```
Db 73 VPRGDFGAGQRCG---PLPGAPASALVDHLQVDDAVPLGLPRLHLAVDDAH----- 122
Oy 131 APOPLADEGPPRSIGGAAP-----SLGALD-LIGPATGCL---LDDGLEGLF 174
Db 123 ----RLREGRNGRAGSGAAPRLRVRAAGGAGHTVAADVLEPVGACGAGVQGRGTG-- 176
Oy 175 EDIDTSMDNDELMAAPSEGLKRP--ED-----GPGKEAPEL--DEAEIDYLM 220
Db 177 ----ARLDREMHITDAEGIGLGPRLQEESEFGSGPLGPIQLTLPLLFEGCNBPVOD 231
Oy 221 VLWG--TQALERRPGP 234
Db 232 VLLGLPSAVLDRPPLP 247

RESULT 8
US-10-091-483-139
; Sequence 139, Application US/10091483
; Publication No. US20030049650A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PFI212C1
; CURRENT APPLICATION NUMBER: US/10/091,483
; NUMBER OF SEQ ID NOS: 348
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 139
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (18)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (19)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (39)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-10-091-483-139

Query Match
Best Local Similarity 26.6%; Score 115; DB 15; Length 255;
Matches 68; Conservative 27; Mismatches 83; Indels 78; Gaps 14;

Oy 28 PGHAUVAQAP-----AVASSSLFLSVLKLNHSLOQSEPDRLHLYLVNT 73
Db 21 PGPGEASALPGDGTGCGGLYSAKAPNFMWRCV-----RRRLSLNPLFLPINA 72
Oy 74 LRR--IQASMAAPALPVPSPAPASVADULASSDA--LSASNASLLEDLSHTEGLSQ 130
Db 73 VPRGDFGAGQRCG---PLPGAPASALVDHLQVDDAVPLGLPRLHLAVDDAH----- 122
Oy 131 APOPLADEGPPRSIGGAAP-----SLGALD-LIGPATGCL---LDDGLEGLF 174
Db 123 ----RLREGRNGRAGSGAAPRLRVRAAGGAGHTVAADVLEPVGACGAGVQGRGTG-- 176
Oy 175 EDIDTSMDNDELMAAPSEGLKRP--ED-----GPGKEAPEL--DEAEIDYLM 220
Db 177 ----ARLDREMHITDAEGIGLGPRLQEESEFGSGPLGPIQLTLPLLFEGCNBPVOD 231
Oy 221 VLWG--TQALERRPGP 234
Db 232 VLLGLPSAVLDRPPLP 247
```

```
RESULT 9
US-09-912-917-2
; Sequence 2, Application US/09912917
; Patent No. US20020028487A1
; GENERAL INFORMATION:
; APPLICANT: Lathangue, Nicholas B
; APPLICANT: Shikama, No. US20020028487A1ko
; TITLE OF INVENTION: The University Court of the University of Glasgow
; FILE REFERENCE: JMY, a co-activator for p300/CBP, nucleic acid encoding
; CURRENT APPLICATION NUMBER: US/09/912,917
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 09/311,236
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: GB 9818235.5
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-912-917-2

Query Match
Best Local Similarity 25.1%; Score 113; DB 9; Length 983;
Matches 61; Conservative 26; Mismatches 84; Indels 72; Gaps 11;

Oy 3 SKGLKRKEEEREKPLAVDSWMLDPGHAUVAQAP-----PAAVSSSLFLSVLKLNHS 56
Db 94 SPGPRKSOAMAEKGSPPRSARSLKDPGPPGAPGEPSELRSPARAKAS----- 141
Oy 57 LQSEPDRLHLYLVNTLRIQASWAPPAALPVPSPAPASVADNLLASDAALSMA 116
Db 142 ----PURAESNDALASATPPRPAPVPVSVRVASGAVS----- 181
Oy 117 SLEDLSHIEGL--SQADQPLAD--EGPPRSIGGAAPSLGALDLGPATGCLDDGLEGL 173
Db 182 --EEIEVLEWREDEADQPLPDSQPP-----SAAELSSAECC--SWAGL 223
Oy 174 FEDIDTSMDNDELMAAPSEGLKRP-----GPDGPKR-----EAPRLDAEID--YLM 221
Db 224 FSPQDLRAVHQQLCSVNSQ--LEPCLPVPPEPSPGMMTVLFGADPEMTQEIADALCYQLQV 282
Oy 222 LVG 224
Db 283 YLG 285

RESULT 10
US-10-157-031-415
; Sequence 415, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequence
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 415
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-415

Query Match
Best Local Similarity 9.3%; Score 112; DB 15; Length 715;
```

[illegible]

```

RESULT 11
US-10-260-715-8
Sequence 8, Application US/10260715
Publication No. US2003009992A1
GENERAL INFORMATION:
APPLICANT: UCB, S.A.
APPLICANT: NOCKA, Karl
APPLICANT: LU, Sun
APPLICANT: MEDLEY, Quintus
APPLICANT: THOMIS, Daniel
APPLICANT: GU, Jesse
TITLE OF INVENTION: Genes Associated with Mast Cell Activation
FILE REFERENCE: 053529-5006
CURRENT APPLICATION NUMBER: US/10/260,715
CURRENT FILING DATE: 2002-10-01
PRIOR APPLICATION NUMBER: 60/325,536
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 1259
TYPE: PR1
ORGANISM: Homo sapiens
US-10-260-715-8

```

| | | | | |
|-----------------------|--------|----------------|--------|----------------|
| Query Match | 9.1%; | Score 110; | DB 15; | Length 1259; |
| Best Local Similarity | 26.2%; | Pred. No. 1.3; | | |
| Matches | 48; | Conservative | 20; | Mismatches 71; |
| | | | | Indels 44; |
| | | | | Gaps 8; |

[illegible]

| | | | |
|----|-----|-----|-----|
| Oy | 232 | PGP | 234 |
| | 111 | | |
| Db | 778 | PGP | 780 |

RESULT 12
US-10-156

Sequence 12958, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO

```

1  APPLICANT:  ISHIKAWA, JUN
2  APPLICANT:  HORIKAWA, HIROSHI
3  APPLICANT:  SHIDA, TADAYOSHI
4  APPLICANT:  SAKAKI, YOSHIYUKI
5  APPLICANT:  HATTORI, MASAHIRA
6  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
7  FILE REFERENCE: 249-262
8  CURRENT APPLICATION NUMBER: US/10/156,761
9  CURRENT FILING DATE: 2002-05-29
10 PRIOR APPLICATION NUMBER: JP 2001-204089
11 PRIOR FILING DATE: 2001-05-30
12 PRIOR APPLICATION NUMBER: JP 2001-272697
13 PRIOR FILING DATE: 2001-08-02
14 NUMBER OF SEQ. ID NOS: 15109
15 SEQ. ID NO 12958
16 LENGTH: 803
17 TYPE: part
18 ORGANISM: Streptomyces avermitilis
19 US-10-156-761-12958

```

| | | | | |
|-----------------------|--------|------------------|----------------|-------------|
| Query Match | 9.0%; | Score 108.5; | DB 15; | Length 803; |
| Best Local Similarity | 25.8%; | Pred. No. 0.97; | | |
| Matches | 49; | Conservative 15; | Mismatches 65; | Indels 61; |
| | | | | Gaps 8 |

QY 28 PGAAAAQ-----AP-PAVAASSLDELVLKHSLSQSEPDRLHLYVNTLEIRIQASM 81
 Db 456 FGHATAVQPSHSPAPAPALPTPLPFP--AQCHHP-----EAPL 4919
 QY 82 AAPAALP---PVPSPAPASVADNLNLSASDALSMSASLLEDSHTEGLSQAPPLADE 138
 Db 492 VPTTASPTASPTASPTASPAVGN---ASTEAPASGRAPLSPGYPERAGACERP----- 542Z
 QY 139 GPFGKSIIGGAAPSLGALLDLPATGCLDDGLEGLEFDITSMYDNELMAPASGLKPGP 1989
 Db 543 -----SRGGAAAPGTGSHALANAAVTASRLPMPG-----APTSSPMSPGE 5633

```

QY      199  EDGPKEAP  208
        ||      ||
Db      584  LPPGALPSAP 593

```

RESULT 13
US-09-963-896-7
; Sequence 7, Application US/09963896
; Patent No. US20020102585A1

APPLICANT: Lal, Preeti
Guegler, Karl J.

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA

ZIP: 94304
COMPUTER READABLE

```

:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
:
: CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/09/963,896
; FILING DATE: 26-Sep-2001
; CLASSIFICATION: <Unknown>
;
PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 09/397,558
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:

NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132

```

; REFERENCE/DOCKET NUMBER: PF-0527 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1317 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 2459993
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-963-896-7

Query Match      8.6%; Score 104; DB 10; Length 1317;
Best Local Similarity 23.7%; Pred. No. 2.4; Mismatches 79; Indels 88; Gaps 10;
Matches 58; Conservative 20;

OY 78 QASMAPAALPPVSP-----PAAPSVADN-----LIASSDA 109
DB 746 EGSAEPQLPSPSPSCGASLPSEASAPDILPASPTPAAGSVTVPEPAPTLSSGS 805
OY 110 ALSASMSLDELDSHIEGLSQAPQPLADEGPPGKSGIC-----GAAPSGALDL 157
DB 806 SLGGEAPG-SHEDDTTEATSGVFTDLSGDPHTKSGIVPALRSLOKQVGTPTDLSIDI 864
OY 158 LGPAT--GC-----LLDDGL-----EGLFEDIDTSMYDNELMAPAS 191
DB 865 PSSASDGCCEVLSPSAGPPGQPRANDSGIDTENTESPVELKFAHSSPEAFGEFAS 924
OY 192 EGLKPGPD-----GPKKEAPELDEA--ELDYLDVIVG-----TQALER 230
DB 925 EGSPGPDPLSVSLGSLSKSPYRDSAYFSDLDASEPPTGPKKHSGIQDSQKEODIRS 984
OY 231 PPGPG 235
DB 985 PPSPG 989

RESULT 14
US-10-076-069-6
; Sequence 6, Application US/10076069
; Publication No. US2002017714A1
; GENERAL INFORMATION:
; APPLICANT: JURERIC, ROLAND
; APPLICANT: NACHTMAN, RONALD
; TITLE OF INVENTION: HEP, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEVELOPMENT
; FILE REFERENCE: 39532-176599
; CURRENT APPLICATION NUMBER: US/10/076, 069
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/268, 923
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-076-069-6

Query Match      8.6%; Score 104; DB 14; Length 113;
Best Local Similarity 34.1%; Pred. No. 0.19; Mismatches 26; Indels 12; Gaps 2;
Matches 28; Conservative 16;

OY 1 MLSGKRRKEEKEEPLAVDSWMLDPAVAQAAPPAVASSSLPDLSTVLKHLHSIQGS 60
DB 1 MFAAGLKRKIGDDE-----GVGEGIVPSTSLQRO-----SLDMSLVKQLDCHMLV 48
OY 61 EPDRLRLVVLVNTLRRIQASMA 82
DB 49 EPNLCRSVLTANTVRAIOIEKMS 70
```

```

RESULT 15
US-09-789-561-97
; Sequence 97, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 97
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (345)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-97

Query Match      8.6%; Score 104; DB 9; Length 459;
Best Local Similarity 24.6%; Pred. No. 1.1; Mismatches 68; Indels 86; Gaps 11;
Matches 56; Conservative 18;

OY 24 WMLDPAHAQAAPPAVASSSLPDLSTVLKHLHSIQSEPDRLHLVL----- 69
DB 93 WRCCPGHSGVSCCEVAASASLSEPM-----WSGSTWRMRALPTAFSCCLNCSKY 142
OY 70 --VNTLRRIQASMAPAAL--PPVSPPAASVADNLLASDALSMSMSLDELDSHIE 126
DB 143 SELTERLKVLEAKMTMLTVIEQVPPTPATP-----EDPAPLV 180
OY 127 GLSQAPQPLADEGPPGKSGIGGAAPSLGALDLGPAFGCLDDGLEGTFEDIDTSMYDNEL 186
DB 181 G-----PPPA--QGSFGD--GGLDQOVGAMGLPGP--TGPKGDAAGSG----- 217
OY 187 WAPASEGKPGEPDEGPKKEAPELDEAELDYLDVIVGTOALERPPGP 234
DB 218 --PMGMRGPPGPGQPGG---SPGRAGA-----VGTPOERGPPEGP 251
```

Search completed: August 19, 2003, 09:44:59
Job time : 58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 15:06:01 ; Search time 310 Seconds

(without alignments)
8644.552 Million cell updates/sec

Title: US-09-701-675a-8

Perfect score: 1192
Sequence: 1 gtagacgcgagtcagactg.....gggaatactcaaaaaaaaaa 1192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

- Published Applications, NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/2/pubpna/PC1_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/2/pubpna/PC1US_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
 - 10: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
 - 12: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
 - 14: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
 - 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 - 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
 - 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|---------|-------|---------------------------------------|
| 1 | 1188.8 | 99.7 | 1219 | 10 | US-09-731-872-171 Sequence 171, App |
| 2 | 1188.8 | 99.7 | 1219 | 12 | US-09-876-997-171 Sequence 171, App |
| 3 | 258 | 21.6 | 258 | 10 | US-09-796-692-5565 Sequence 5565, App |
| 4 | 258 | 21.6 | 258 | 14 | US-10-040-862-5565 Sequence 5565, App |
| 5 | 120.8 | 10.1 | 165 | 10 | US-09-728-445-520 Sequence 520, App |
| 6 | 46.2 | 3.9 | 1227 | 14 | US-10-156-761-2641 Sequence 2641, App |
| 7 | 46.2 | 3.9 | 9025608 | 14 | US-10-156-761-1 Sequence 1, Appl1 |
| 8 | 44.6 | 3.7 | 2082 | 13 | US-10-076-069-3 Sequence 3, Appl1 |
| 9 | 44.6 | 3.7 | 2629 | 9 | US-10-037-270-773 Sequence 773, App |
| 10 | 43 | 3.6 | 396 | 9 | US-09-825-294-167 Sequence 167, App |
| 11 | 43 | 3.6 | 396 | 10 | US-09-970-966-167 Sequence 167, App |
| 12 | 43 | 3.6 | 396 | 15 | US-10-212-677-167 Sequence 167, App |
| 13 | 43 | 3.6 | 2038 | 9 | US-09-925-302-153 Sequence 153, App |
| 14 | 43 | 3.6 | 2038 | 14 | US-10-106-698-179 Sequence 179, App |
| 15 | 43 | 3.6 | 9025608 | 14 | US-10-156-761-1 Sequence 1, Appl1 |
| 16 | 42.2 | 3.5 | 1049 | 12 | US-10-140-472-358 Sequence 358, App |

| | | | | | | |
|------|------|-----|-------|----|----------------------|----------------------|
| c 17 | 42.2 | 3.5 | 1049 | 12 | US-10-141-761-358 | Sequence 358, App |
| c 18 | 42.2 | 3.5 | 1049 | 14 | US-10-123-155-358 | Sequence 358, App |
| c 19 | 42.2 | 3.5 | 1049 | 15 | US-10-146-731-358 | Sequence 358, App |
| 20 | 42.2 | 3.5 | 1428 | 14 | US-10-145-415-15 | Sequence 15, Appl |
| 21 | 42.2 | 3.5 | 2082 | 13 | US-10-076-069-1 | Sequence 1, Appl1 |
| 22 | 42 | 3.5 | 10056 | 14 | US-10-156-761-411 | Sequence 411, App |
| 23 | 42 | 3.5 | 10000 | 14 | US-10-156-761-15103 | Sequence 15103, A |
| 24 | 41.8 | 3.5 | 1413 | 14 | US-10-145-415-5 | Sequence 5, Appl1 |
| 25 | 41.8 | 3.5 | 1683 | 14 | US-10-156-761-5824 | Sequence 5824, App |
| c 26 | 41.4 | 3.5 | 2949 | 14 | US-10-187-267A-12 | Sequence 12, Appl |
| c 27 | 41.4 | 3.5 | 36321 | 11 | US-10-187-267A-1 | Sequence 1, Appl1 |
| c 28 | 40.8 | 3.4 | 1387 | 14 | US-09-756-841C-1 | Sequence 1, Appl1 |
| c 29 | 40.6 | 3.4 | 378 | 14 | US-10-156-761-332 | Sequence 332, App |
| c 30 | 40.6 | 3.4 | 474 | 11 | US-09-918-995-3967 | Sequence 3967, App |
| c 31 | 40.4 | 3.4 | 2940 | 14 | US-10-156-761-3172 | Sequence 3172, App |
| c 32 | 40.2 | 3.4 | 268 | 11 | US-09-373-658-58 | Sequence 58, Appl |
| c 33 | 39.4 | 3.3 | 1344 | 14 | US-10-156-761-1108 | Sequence 1108, App |
| c 34 | 39 | 3.3 | 1836 | 14 | US-10-156-761-2859 | Sequence 2859, App |
| c 35 | 39 | 3.3 | 13862 | 11 | US-09-764-891-5477 | Sequence 5477, App |
| c 36 | 39 | 3.3 | 13862 | 11 | US-09-764-891-10204 | Sequence 10204, A |
| c 37 | 39 | 3.3 | 13862 | 14 | US-10-205-428-1003 | Sequence 1003, App |
| c 38 | 38.8 | 3.3 | 735 | 13 | US-10-027-632-158772 | Sequence 158772, App |
| c 39 | 38.8 | 3.3 | 31871 | 10 | US-09-764-847-1403 | Sequence 1403, App |
| c 40 | 38.8 | 3.3 | 31871 | 14 | US-10-092-154-1403 | Sequence 1403, App |
| c 41 | 38.4 | 3.2 | 2109 | 10 | US-09-738-973-153 | Sequence 153, App |
| c 42 | 38.4 | 3.2 | 2109 | 14 | US-09-854-133-153 | Sequence 153, App |
| c 43 | 38.4 | 3.2 | 2109 | 14 | US-10-144-649A-153 | Sequence 153, App |
| c 44 | 38.4 | 3.2 | 3377 | 13 | US-10-027-632-113850 | Sequence 113850, App |
| c 45 | 38.2 | 3.2 | 2968 | 9 | US-09-815-915-1 | Sequence 1, Appl1 |

ALIGNMENTS

RESULT 1
US-09-731-872-171
Sequence 171, Application US/09731872
Patent No. US20020102604A1
GENERAL INFORMATION:
APPLICANT: Dumas Maline Edwards, Jean Baptiste
APPLICANT: Bougueleret, Lydie
APPLICANT: Jobert, Severin
TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
FILE REFERENCE: 78. US3. REG
CURRENT APPLICATION NUMBER: US/09/731,872
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,629
PRIOR FILING DATE: 1999-12-08
PRIOR APPLICATION NUMBER: US 60/187,470
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 482
SOFTWARE: Patent.pm
SEQ ID NO 171
LENGTH: 1219
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 160..867
US-09-731-872-171

Query Match 99.7%; Score 1188.8; DB 10; Length 1219;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

| | | | |
|----|----|--|-----|
| QY | 1 | GTGACCGGAGTCAAACTGCTCTCGGAGCAGCCAGCGGGGTTCCGGAGCAGCCAA | 60 |
| DB | 21 | GTGACCGGAGTCAAACTGCTCTCGGAGCAGCCAGCGGGGTTCCGGAGCAGCCAA | 80 |
| QY | 61 | CAAGGAGTCTGCGCGCGCGGTTCTGTTGTTGGTGGTACCTCTTCTGTTGTA | 120 |
| DB | 81 | CAAGGAGTCTGCGCGCGCGGTTCTGTTGTTGGTGGTACCTCTTCTGTTGTA | 140 |

121 TTGCGCCCTAGTAGAGCAAGATGCTGAGCAAGGGCTCTGAAGCGGAAACGGAGAGAGAGA 180
141 TTGCGCCCTAGTAGAGCAAGATGCTGAGCAAGGGCTCTGAAGCGGAAACGGAGAGAGAGA 200
181 GGAGAGAGAGACCTCTGGGAGAGTGCAGTCTCTGGGAGTATATCTGGCCACGACGGGTGGC 240
201 GGAGAGAGAGACCTCTGGGAGAGTGCAGTCTCTGGGAGTATATCTGGCCACGACGGGTGGC 260
241 ACAGGACACCCCGGCGCTGAGCTAGCTCCCTCTTTAGACCTCTAGAGCTCAAGCTCA 300
261 ACAGGACACCCCGGCGCTGAGCTAGCTCCCTCTTTAGACCTCTAGAGCTCAAGCTCA 320
301 CCACAGCTGAGAGAGAGTGAAGCGGAGCTGGCGGACCTGGTGTCTGAGTCAAGCTCT 360
321 CCACAGCTGAGAGAGTGAAGCGGAGCTGGCGGACCTGGTGTCTGAGTCAAGCTCT 380
361 GCGGGGCAATCCAGGGGTGATGGGACCGGCGCTGCTGGCACCTGGTGTCTGAGTCAAGCTCT 420
381 GCGGGGCAATCCAGGGGTGATGGGACCGGCGCTGCTGGCACCTGGTGTCTGAGTCAAGCTCT 440
421 TGACACCCCGGCGCTGAGCTAGCTCCCTCTTTAGACCTCTAGAGCTCAAGCTCA 480
441 TGACACCCCGGCGCTGAGCTAGCTCCCTCTTTAGACCTCTAGAGCTCAAGCTCA 500
481 CATGGCCAGCTCTGAGAGAGCTGAGCAACATTGAGGAGCTGAGTCAAGCTCCCAACC 540
501 CATGGCCAGCTCTGAGAGAGCTGAGCAACATTGAGGAGCTGAGTCAAGCTCCCAACC 560
541 CTGGCAGACGAGGGGGCCACGAGCGCTGAGATCGGGGAGAGCGGCGGACCTGGGTC 600
561 CTGGCAGACGAGGGGGCCACGAGCGCTGAGATCGGGGAGAGCGGCGGACCTGGGTC 620
601 CTGGACCTGCTGGGCGCCAGCCACTGCTGTCTACTGAGCAGATGGGCTTGAAGGCTCTTT 660
621 CTGGACCTGCTGGGCGCCAGCCACTGCTGTCTACTGAGCAGATGGGCTTGAAGGCTCTTT 680
661 TGAGGATTTGACACCTCTATGATGACAAATGAACTTTGGGACACAGCTCTGAGGCTCT 720
681 TGAGGATTTGACACCTCTATGATGACAAATGAACTTTGGGACACAGCTCTGAGGCTCT 740
721 CAACACGAGCCCTGAGAGTGGGCGGAGAGAGAGAGCTCCGAGCTGAGAGAGAGAGAGA 780
741 CAACACGAGCCCTGAGAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 800
781 ATTGACTACTCATGATGATGCTGCTGGGAGACACAGGCACTGAGAGAGAGAGAGAGAGAG 840
801 ATTGACTACTCATGATGATGCTGCTGGGAGACACAGGCACTGAGAGAGAGAGAGAGAGAG 860
841 AGGGGCTGAGAGCCCTGCTGAG 900
861 AGGGGCTGAGAGCCCTGCTGAG 920
901 GAGCAACTGCTCTGAG 960
921 GAGCAACTGCTCTGAG 980
961 CACTTTGAG 1020
981 CACTTTGAG 1040
1021 CTGGCAGGAGGAG 1080
1041 CTGGCAGGAGGAG 1100
1081 GAATTCCTGAG 1140
1101 TAATTCCTGAG 1160
1141 CCAATTCAGATTTTCAATTAAG 1192
1161 CCAATTCAGATTTTCAATTAAG 1212

RESULT 2
US-09-876-997-171
; Sequence 171, Application US/09876997
; Publication No. US20030152921A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bouqueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78 USA, CIP
; CURRENT APPLICATION NUMBER: US/09/876, 997
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 09/731, 872
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/187, 470
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: US 60/169, 629
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 171
; LENGTH: 1219
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 160..867
US-09-876-997-171
Query Match 99.7%; Score 1188.8; DB 12; Length 1219;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 GTGAGCCGAGATCGAAGATGCTGCTCGGAGCCAGGCGGGGTTCCGAGAGAGAGCA 60
21 GTGAGCCGAGATCGAAGATGCTGCTCGGAGCCAGGCGGGGTTCCGAGAGAGAGCA 80
61 CAAGGATGCTGCGCGCCGCTTCCTGATTTGTTGGGAGGCTGACTCTTCCTTGA 120
81 CAAGGATGCTGCGCGCCGCTTCCTGATTTGTTGGGAGGCTGACTCTTCCTTGA 140
121 TTGGCCCTGATGAGCAAGATGCTGAGCAAGAGGCTGTAAGCGGAAAGGAGAGAGA 180
141 TTGGCCCTGATGAGCAAGATGCTGAGCAAGAGGCTGTAAGCGGAAAGGAGAGAGA 200
181 GGAGAGAGAGACCTCTGGGAGAGTGCAGTCTCTGGGAGTATATCTGGCCACGACGGGTGGC 240
201 GGAGAGAGAGACCTCTGGGAGAGTGCAGTCTCTGGGAGTATATCTGGCCACGACGGGTGGC 260
241 ACAGGACACCCCGGCGCTGAGCTAGCTCCCTCTTTAGACCTCTAGAGCTCAAGCTCA 300
261 ACAGGACACCCCGGCGCTGAGCTAGCTCCCTCTTTAGACCTCTAGAGCTCAAGCTCA 320
301 CCACAGCTGAGAGAGAGTGAAGCGGAGCTGGCGGACCTGGTGTCTGAGTCAAGCTCT 360
321 CCACAGCTGAGAGAGTGAAGCGGAGCTGGCGGACCTGGTGTCTGAGTCAAGCTCT 380
361 GCGGCGATCCAGGCGCTGATGGGACCGGCGCTGCTGGCACCTGGTGTCTGAGTCAAGCTCT 420
381 GCGGCGATCCAGGCGCTGATGGGACCGGCGCTGCTGGCACCTGGTGTCTGAGTCAAGCTCT 440
421 TGACACCCCGGCGCTGAGCTAGCTCCCTCTTTAGACCTCTAGAGCTCAAGCTCA 480
441 TGACACCCCGGCGCTGAGCTAGCTCCCTCTTTAGACCTCTAGAGCTCAAGCTCA 500
481 CATGGCCAGCTCTGAGAGAGCTGAGCAACATTGAGGAGCTGAGTCAAGCTCCCAACC 540
501 CATGGCCAGCTCTGAGAGAGCTGAGCAACATTGAGGAGCTGAGTCAAGCTCCCAACC 560
541 CTGGCAGACGAGGGGGCCACGAGCGCTGAGATCGGGGAGAGCGGCGGACCTGGGTC 600
561 CTGGCAGACGAGGGGGCCACGAGCGCTGAGATCGGGGAGAGCGGCGGACCTGGGTC 620

| | | | |
|----|------|---|------|
| OY | 601 | CTTGAGACTGCTGGGCCAGCAGCATGGCTGTCTACTGACATATGGCTTGAGGGCTGT | 660 |
| Db | 621 | CTTGAGACTGCTGGGCCAGCAGCATGGCTGTCTACTGACATATGGCTTGAGGGCTGT | 680 |
| OY | 661 | TGAGGATATTTGACACCTCTATGTATGACAAATGAACCTTTGGGCACACGCTCTGAGGGCCT | 720 |
| Db | 681 | TGAGGATATTTGACACCTCTATGTATGACAAATGAACCTTTGGGCACACGCTCTGAGGGCCT | 740 |
| OY | 721 | CAAAACGAGGCTCTAGAGATGGGCCGGGCGMAGGAGAAAGCTCCGAGCTGGACAGGCCGA | 780 |
| Db | 741 | CAAAACGAGGCTCTAGAGATGGGCCGGGCGMAGGAGAAAGCTCCGAGCTGGACAGGCCGA | 800 |
| OY | 781 | ATTGAGCTACCTCAATNGATGTGTGCTGGTGGGCACACAGGACACTGGAGGACCGCGGGGGCC | 840 |
| Db | 801 | ATTGAGCTACCTCAATNGATGTGTGCTGGTGGGCACACAGGACACTGGAGGACCGCGGGGGCC | 860 |
| OY | 841 | AGGGCGCTGAGCCCTCGTGTGTAATGGTTGTCTGATCTGAACTGAGCTGCTGGCTG | 900 |
| Db | 861 | AGGGCGCTGAGCCCTCGTGTGTAATGGTTGTCTGATCTGAACTGAGCTGCTGGCTG | 920 |
| OY | 901 | GACCACTGTCTCTGAAAAGACACAGCTGGCTTCCCTAGTACAGAGAACAGAGGCTTGGGC | 960 |
| Db | 921 | GACCACTGTCTCTGAAAAGACACAGCTGGCTTCCCTAGTACAGAGAACAGAGGCTTGGGC | 980 |
| OY | 961 | CACCTTGGAGAGACAGAACTAGTACGTGGGGCAATTACATCCGCTCCCTCGTCTCAGGG | 1020 |
| Db | 981 | CACCTTGGAGAGACAGAACTAGTACGTGGGGCAATTACATCCGCTCCCTCGTCTCAGGG | 1040 |
| OY | 1021 | CTGGCAGGGGAGCCTGGAAATTACCCCTAGTGTGATGTGATGACAGGGTCTGGTGGGACT | 1080 |
| Db | 1041 | CTGGCAGGGGAGCCTGGAAATTACCCCTAGTGTGATGTGATGACAGGGTCTGGTGGGACT | 1100 |
| OY | 1081 | GAATTCCTGGCCCTGGGGGTCAATAGCTTGGGCTGTCCCTCTCTGTGATACGGGAAGAGACC | 1140 |
| Db | 1101 | TAAATTCCTGGCCCTGGGGGTCAATAGCTTGGGCTGTCCCTCTCTGTGATACGGGAAGAGACC | 1160 |
| OY | 1141 | CCAAATCGAATTTTCAAAATTAACCCAGCTCTGGGAAATTCACAAAAAATAA | 1192 |
| Db | 1161 | CCAAATCGAATTTTCAAAATTAACCCAGCTCTGGGAAATTCACAAAAAATAA | 1212 |

RESULT 3
US-09-796-692-5565/c
: Sequence 5565, Application US/09796692
: Publication No. US20020198362A1
: GENERAL INFORMATION:
: APPLICANT: Gaiger, Alexander
: APPLICANT: Algate, Paul A.
: APPLICANT: Mannion, Jane
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
: TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
: FILE REFERENCE: 2077_001200
: CURRENT APPLICATION NUMBER: US/09/796,692
: CURRENT FILING DATE: 2001-03-01
: PRIOR APPLICATION NUMBER: 60/186,126
: PRIOR FILING DATE: 2000-03-01
: PRIOR APPLICATION NUMBER: 60/190,479
: PRIOR FILING DATE: 2000-03-17
: PRIOR APPLICATION NUMBER: 60/200,545
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: 60/200,303
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,779
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,999
: PRIOR FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: 60/202,084
: PRIOR FILING DATE: 2000-05-04
: PRIOR APPLICATION NUMBER: 60/206,201
: PRIOR FILING DATE: 2000-05-22
: PRIOR APPLICATION NUMBER: 60/218,950
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: 60/222,903

```

: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: 60/223,416
: PRIOR FILING DATE: 2000-08-04
: PRIOR APPLICATION NUMBER: 60/223,378
: PRIOR FILING DATE: 2000-08-07
: NUMBER OF SEQ ID NOS: 9597
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5565
: LENGTH: 258
: TYPE: DNA
: ORGANISM: Homo sapiens
US-03-796-692-5565

```

| | Query Match | Similarity | 21.6% | Score 258 | DB 10 | Length 258 |
|----|-------------|---|--------------|-----------------|------------|------------|
| | Best Local | Similarity | 100.0% | Pred. No. 5e-66 | | |
| | Matches | 258 | Conservative | 0 | Mismatches | 0 |
| | | | | | Indels | 0 |
| | | | | | Gaps | 0 |
| QY | 927 | CTGGCTTCCCTAGTACAGAGAACAAGGCTTGAGGCACCTTGGAGACAGACAATCAGTCC | 986 | | | |
| | | | | | | |
| DB | 258 | CTGGCTTCCCTAGTACAGAGAACAAGGCTTGAGGCACCTTGGAGACAGACAATCAGTCC | 199 | | | |
| QY | 987 | TGGGCACTTCACATCCGTCCCTCTGTCTCAGGGCTGGCAGGGGAGCCTGGAATTACC | 1046 | | | |
| | | | | | | |
| DB | 198 | TGGGCACTTCACATCCGTCCCTCTGTCTCAGGGCTGGCAGGGGAGCCTGGAATTACC | 139 | | | |
| QY | 1047 | CCTAGTATGGAATGACAGGGCTGTGGTGGGAGCTGAATTCCTCGGCGGTATAGC | 1106 | | | |
| | | | | | | |
| DB | 138 | CCTAGTATGGAATGACAGGGCTGTGGTGGGAGCTGAATTCCTCGGCGGTATAGC | 79 | | | |
| QY | 1107 | TTGGGCTTCTCTCTGTATACGGGAGAGACCCCATTCAGATTTTCAATTTAAACC | 1166 | | | |
| | | | | | | |
| DB | 78 | TTGGGCTTCTCTCTGTATACGGGAGAGACCCCATTCAGATTTTCAATTTAAACC | 19 | | | |
| QY | 1167 | AGTCCTGGGAATCTCAA | 1184 | | | |
| | | | | | | |
| DB | 18 | AGTCCTGGGAATCTCAA | 1 | | | |

```

1  RESULT 4
2  US-10-040-862-5565/c
3  Sequence 5565, Application US/10040862
4  Publication No. US20030078396A1
5  GENERAL INFORMATION:
6  APPLICANT: Gaiger, Alexander
7  APPLICANT: Algate, Paul A.
8  APPLICANT: Mannion, Jane
9  APPLICANT: Reltter, Marc
10 APPLICANT: Corixa Corporation
11 TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
12 TITLE OF INVENTION: Hemtological Malignancies
13 FILE REFERENCE: 014058-013520US
14 CURRENT APPLICATION NUMBER: US/10/040, 862
15 CURRENT FILING DATE: 2001-11-06
16 PRIOR APPLICATION NUMBER: US 60/186,126
17 PRIOR FILING DATE: 2000-03-01
18 PRIOR APPLICATION NUMBER: US 60/190,479
19 PRIOR FILING DATE: 2000-03-17
20 PRIOR APPLICATION NUMBER: US 60/200,545
21 PRIOR FILING DATE: 2000-04-27
22 PRIOR APPLICATION NUMBER: US 60/200,303
23 PRIOR FILING DATE: 2000-04-28
24 PRIOR APPLICATION NUMBER: US 60/200,779
25 PRIOR FILING DATE: 2000-04-28
26 PRIOR APPLICATION NUMBER: US 60/200,999
27 PRIOR FILING DATE: 2000-05-01
28 PRIOR APPLICATION NUMBER: US 60/202,084
29 PRIOR FILING DATE: 2000-05-04
30 PRIOR APPLICATION NUMBER: US 60/206,201
31 PRIOR FILING DATE: 2000-05-22
32 PRIOR APPLICATION NUMBER: US 60/218,950
33 PRIOR FILING DATE: 2000-07-14
34 PRIOR APPLICATION NUMBER: US 60/222,903
35 PRIOR FILING DATE: 2000-08-03

```

RESULT 7
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match
Best Local Similarity 45.3%; Score 46.2; DB 14; Length 9025608;
Matches 168; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 483 TGCCACAGCTCTGTGAGGACCTTCAGCCACATTGAGGGCTTGAGGCTCCCAACCT 542
DB 3253368 TCGTCCGCGCCCTCGCGGCGCGATGATTGACGGCACCCCGACGCTGCGCAGC 3253427
QY 543 TGGCAGACGAGGGGCGCACCGCGCGTACGATTCGGGGAGCAGCGCCGACGCTGGCT 602
DB 3253428 GCGATGACGATGCGCGCCCTTGGGCGGTGCGCTCCAGCTTGATGACGATGCGGTGATA 3253487
QY 603 TGGACCTGCTGGGCGCCAGCCACGACCTGCTTACTGACGATGGGCTTGAAGGCTTTTG 662
DB 3253488 TTGACGACCTCGCGCGAAGACACTGCTGACACGCGCTTTCGCCCGGTGGGCTCG 3253547
QY 663 AGGATATTGACACCTGTATGTATGACATGAACCTTTGGGACACCGCTTGAGGGCTCA 722
DB 3253548 AGGACGACGATGACTCGTCGACGGGGGCTGCTTTCGACGACGGGCTTGACCTTGCCC 3253607
QY 723 AACCAAGCCCTGAGGATGGGCGCGGCAAGAGAGACTCCGGAGCTGGACGAGCGCGAAT 782
DB 3253608 AGCTGCTCCATGAGCGCGGCTTGGTGTGACGCGCGCCGCGGTGTCGATGAGAGCAGC 3253667
QY 783 TGGACTAACCTGATGATGCTGCTGGGACACACAGGACACTGAGCGACCGCGGGGCGAG 842
DB 3253668 TCGGCGCCCTCCCTCGCATGCCCTCTTGACGGGCTCGAAGGCGATTCACGCGGGTGC 3253727
QY 843 GCGCGTGAGGC 853
DB 3253728 CCCTCGGGTCC 3253738

RESULT 8
US-10-076-069-3
Sequence 3, Application US/10076069
Publication No. US20020177214A1
GENERAL INFORMATION:
APPLICANT: JURECIC, ROLAND
APPLICANT: NACHTMAN, RONALD
TITLE OF INVENTION: HEPY, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEVELOPMENT
FILE REFERENCE: 39532-176599
CURRENT APPLICATION NUMBER: US/10/076,069
CURRENT FILING DATE: 2002-02-15
PRIOR APPLICATION NUMBER: US 60/268,923
PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 2082
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (75)..(797)
US-10-076-069-3

Query Match
Best Local Similarity 50.7%; Score 44.6; DB 13; Length 2082;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 269 TCCCTTTGACCTCAGTGTCAAGCTCCACACAGCCTGACAGAGGTGACCGGAC 328
DB 180 TCGCTCTGAGCATGTCTCTGTGAAGTTGACAGCTTGGCACTTCCTTGTGGAGCCCAAC 239
QY 329 CTGCGCACCTGTGCTGTGTGTAACACTCTGCGGCGCATCCAGGCGTCATGCGACCC 388
DB 240 CTGTGCGCGTCACTCTCATTTGCCAACAGCGTCCGGAGATCCAAAGAGAGATGACGAC 299
QY 389 GCGGCTGCCCTGCCACCTGTGCTTAGCCCACTGACGCCCACTGCTGCTGACAACTTA 448
DB 300 GATGGGACGTGGCGGACAGTGGCAGCCCAAGCTGACAGAGCGGCGGCTGACCGCTTG 359
QY 449 CTGCAAGCTCGGAGCGTGCCCTTTCAGCCT 479
DB 360 GTCTCCACGAGATCTCTGTGCTGCGAGCT 390

RESULT 9
US-10-037-270-773
Sequence 773, Application US/10037270
Publication No. US20030104529A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyang
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aiding J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Tillinghast, John
APPLICANT: Dmanac, Radoje T.
TITLE OF INVENTION: NO. US20030104529A1 Nucleic Acids and
FILE REFERENCE: 784CIP28
CURRENT APPLICATION NUMBER: US/10/037,270
CURRENT FILING DATE: 2002-01-04
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: PL_Fl_genes Version 1.0
SEQ ID NO 773
LENGTH: 2629
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (73)..(798)
US-10-037-270-773

Query Match 3.7%; Score 44.6; DB 14; Length 2629;
Best Local Similarity 50.7%; Pred. No. 0.0077;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

269 TCCCTCTTTGACCTCTCAGTGTCTCAAGCTCCACACAGCTTGACAGAGTGAAGCCGAGC 328
178 TCGCTCTGAGACATGTCTCTGTGTGAAGTTGACGCTTTGGCACATGTCTGTGAAGCCAAAC 237
329 CTGGCGGACCTGTGTCTGTGTGAACACTCTGCGGCGCATCCAGCCCTTCATGCGACCC 388
238 CTGTGCGCGCTCAGTCTCTATTGTCACACAGGTCGCGAGATCCCAAGAGAGATGACGAG 297
389 GCGGCTGCGCTGCGACCTGTGTCTGCTCAAGCTCCACACAGCTTGACAGAGTGTGAGTGAACAATT 448
298 GATGGAGAGTGGCGGACAGTGGCACACCCAGGCTTGACAGAGCGGGCGCGCTGCAGCGCTTG 357
449 CTGGCAAGCTGCGAGCTGCGCTTTTCAGCCT 479
358 GTCTCCACGAGATCTCTGTGCGGTGACGCT 388

Db

RESULT 10
US-09-825-294-167
; Sequence 167, Application US/09825294
; Patent No. US20020004491A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; APPLICANT: Algate, Paul A.
; APPLICANT: Filing, Steven P.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C5
; CURRENT APPLICATION NUMBER: US/09/825,294
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 167
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-825-294-167

Query Match 3.6%; Score 43; DB 9; Length 396;
Best Local Similarity 50.2%; Pred. No. 0.013;
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

269 TCCCTCTTTGACCTCTCAGTGTCTCAAGCTCCACACAGCTTGACAGAGTGAAGCCGAGC 328
173 TCGCTCTGAGACATGTCTCTGTGTGAAGTTGACGCTTTGGCACATGTCTGTGAAGCCAAAT 232
329 CTGGCGGACCTGTGTCTGTGTGAACACTCTGCGGCGCATCCAGCCCTTCATGCGACCC 388
233 CTGTGCGCGCTCAGTCTCTATTGTCACACAGGTCGCGAGATCCCAAGAGAGATGACGAG 292
389 GCGGCTGCGCTGCGACCTGTGTCTGCTCAAGCTCCACACAGCTTGACAGAGTGTGAGTGAACAATT 448
293 GATGGAGAGTGGCGGACAGTGGCACACCCAGGCTTGACAGAGCGGGCGCGCTGCAGCGCTTG 352
449 CTGGCAAGCTGCGAGCTGCGCTTTTCAGCCT 479
353 GTCTCCACGAGATCTCTGTGCGGTGACGCT 383

Db

RESULT 11
US-09-970-966-167
; Sequence 167, Application US/09970966
; Patent No. US20020173638A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Molesh, David Alan
; APPLICANT: Filing, Steven P.

APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C6
; CURRENT APPLICATION NUMBER: US/09/970,966
; CURRENT FILING DATE: 2001-10-02
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-970-966-167

Query Match 3.6%; Score 43; DB 10; Length 396;
Best Local Similarity 50.2%; Pred. No. 0.013;
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

269 TCCCTCTTTGACCTCTCAGTGTCTCAAGCTCCACACAGCTTGACAGAGTGAAGCCGAGC 328
173 TCGCTCTGAGACATGTCTCTGTGTGAAGTTGACGCTTTGGCACATGTCTGTGAAGCCAAAT 232
329 CTGGCGGACCTGTGTCTGTGTGAACACTCTGCGGCGCATCCAGCCCTTCATGCGACCC 388
233 CTGTGCGCGCTCAGTCTCTATTGTCACACAGGTCGCGAGATCCCAAGAGAGATGACGAG 292
389 GCGGCTGCGCTGCGACCTGTGTCTGCTCAAGCTCCACACAGCTTGACAGAGTGTGAGTGAACAATT 448
293 GATGGAGAGTGGCGGACAGTGGCACACCCAGGCTTGACAGAGCGGGCGCGCTGCAGCGCTTG 352
449 CTGGCAAGCTGCGAGCTGCGCTTTTCAGCCT 479
353 GTCTCCACGAGATCTCTGTGCGGTGACGCT 383

Db

RESULT 12
US-10-212-677-167
; Sequence 167, Application US/10212677
; Publication No. US20030129192A1
; GENERAL INFORMATION:
; APPLICANT: Chemault, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Fanger, Gary R.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C7
; CURRENT APPLICATION NUMBER: US/10/212,677
; CURRENT FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 288
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-212-677-167

Query Match 3.6%; Score 43; DB 15; Length 396;
Best Local Similarity 50.2%; Pred. No. 0.013;
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

269 TCCCTCTTTGACCTCTCAGTGTCTCAAGCTCCACACAGCTTGACAGAGTGAAGCCGAGC 328
173 TCGCTCTGAGACATGTCTCTGTGTGAAGTTGACGCTTTGGCACATGTCTGTGAAGCCAAAT 232
329 CTGGCGGACCTGTGTCTGTGTGAACACTCTGCGGCGCATCCAGCCCTTCATGCGACCC 388
233 CTGTGCGCGCTCAGTCTCTATTGTCACACAGGTCGCGAGATCCCAAGAGAGATGACGAG 292
389 GCGGCTGCGCTGCGACCTGTGTCTGCTCAAGCTCCACACAGCTTGACAGAGTGTGAGTGAACAATT 448
293 GATGGAGAGTGGCGGACAGTGGCACACCCAGGCTTGACAGAGCGGGCGCGCTGCAGCGCTTG 352

Db


```

: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 1
: LENGTH: 9025608
: TYPE: DNA
: ORGANISM: Streptomyces avermitilis
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (4187715)
: OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

| | | | | |
|-----------------------|--------------|-----------|---------------|----------------|
| Query Match | 3.6% | Score 43 | DB 14 | length 9025608 |
| Best Local Similarity | 51.6% | Pred. No. | 0.25 | |
| Matches 98 | Conservative | 0 | Mismatches 92 | Indels 0 |
| | | | | Gaps 0 |

| | | | |
|----|---------|--|---------|
| QY | 548 | GACACAGGGGACACAGGACCCGATGATCGGGGGACAGGCCACCTGGGGCTTGGAC | 607 |
| Db | 7420376 | GACCCCGGGAGACCGCTCTCGAAGCCTTGGGGCTCAGCTCTCCGCGCCGGCGCTCGCGAC | 7420317 |
| QY | 608 | CTGCTGGGGCCAGCCACTGGCTGTCTACTGAGACATGGGCTTGAAGGCGTGTTTGAGGAT | 667 |
| Db | 7420316 | GTGGTGCACCGGGGTGCGTGGTCTTGCTGCTGGCCGAGGCCCATCGCGGGCATGATGCCGCC | 7420257 |
| QY | 668 | ATTATCACCTCTATGTATGACATGAACCTTTGGGCACCAAGCCTTGAAGGCGCTTAACA | 727 |
| Db | 7420256 | GACACACACCGCCGAGGACGAGATGACCATGCCCGCCAGAAAGCCAGCGGGTTGGCCAA | 7420157 |
| QY | 728 | GGCCCTGAGG | 737 |
| Db | 7420196 | CACCATGAG | 7420187 |

Search completed: August 25, 2003, 16:56:32
Job time : 334 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 13:39:16 ; Search time 2460 Seconds
(without alignments)
11776.803 Million cell updates/sec

Title: US-09-701-675A-8

Perfect score: 1192
Sequence: 1 gtagccgagatcagaactg.....gggaattcaaaaaaaaaa 1192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrl:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pug:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 856.6 | 71.9 | 1023 | 12 | BM923629 AGENCOURT |
| 2 | 798.8 | 67.0 | 908 | 13 | B0856974 AGENCOURT |
| 3 | 786.6 | 66.0 | 1062 | 12 | BM558458 AGENCOURT |
| 4 | 781.4 | 65.6 | 797 | 13 | BX101521 BX101521 |

| | | | | | | |
|----|-------|------|------|----|----------|---------------------|
| 5 | 756.4 | 63.5 | 766 | 13 | B0539198 | B0539198 AGENCOURT |
| 6 | 751.8 | 63.1 | 803 | 10 | B6715050 | B6715050 602673273 |
| 7 | 749.2 | 62.9 | 868 | 13 | B0433568 | B0433568 AGENCOURT |
| 8 | 744.2 | 62.4 | 883 | 13 | B0878247 | B0878247 AGENCOURT |
| 9 | 743.4 | 62.4 | 765 | 13 | B0537987 | B0537987 AGENCOURT |
| 10 | 742.2 | 62.3 | 782 | 10 | B6468137 | B6468137 602509712 |
| 11 | 740.8 | 62.1 | 846 | 10 | B6829426 | B6829426 602763713 |
| 12 | 737 | 61.8 | 747 | 14 | CD364905 | CD364905 UI-H-FRT2 |
| 13 | 733.2 | 61.5 | 1037 | 12 | B6402912 | B6402912 602418715 |
| 14 | 730.8 | 61.3 | 764 | 13 | B0538362 | B0538362 AGENCOURT |
| 15 | 727 | 61.0 | 949 | 13 | B0957431 | B0957431 AGENCOURT |
| 16 | 722.4 | 60.6 | 923 | 13 | B0957131 | B0957131 AGENCOURT |
| 17 | 716.6 | 60.1 | 732 | 14 | CA748763 | CA748763 UI-H-FRT1 |
| 18 | 708.8 | 59.5 | 878 | 12 | B0690528 | B0690528 AGENCOURT |
| 19 | 699.2 | 58.7 | 782 | 13 | B1753849 | B1753849 603027530 |
| 20 | 693.8 | 58.2 | 876 | 10 | B6898247 | B6898247 601681014 |
| 21 | 693.2 | 58.2 | 882 | 13 | B0690477 | B0690477 AGENCOURT |
| 22 | 687.4 | 57.7 | 697 | 13 | B0182938 | B0182938 UI-H-ED1- |
| 23 | 683.6 | 57.3 | 952 | 13 | B0957038 | B0957038 AGENCOURT |
| 24 | 682.2 | 57.2 | 768 | 10 | B6546883 | B6546883 602574065 |
| 25 | 677 | 56.8 | 698 | 13 | B0633212 | B0633212 UI-H-FRT1- |
| 26 | 662.8 | 55.6 | 695 | 12 | B0973302 | B0973302 UI-CF-EC1 |
| 27 | 656 | 55.0 | 701 | 10 | B6613164 | B6613164 601452139 |
| 28 | 656 | 55.0 | 1189 | 11 | AK004022 | AK004022 Mus muscu |
| 29 | 655.4 | 55.0 | 666 | 14 | CA420820 | CA420820 UI-H-FRT0- |
| 30 | 654.8 | 54.9 | 677 | 14 | CA431558 | CA431558 UI-H-FRT1- |
| 31 | 654.4 | 54.9 | 1185 | 11 | AK008303 | AK008303 Mus muscu |
| 32 | 648 | 54.4 | 656 | 14 | CB851800 | CB851800 UI-CF-FRT0 |
| 33 | 646 | 54.2 | 662 | 14 | CB250543 | CB250543 UI-CF-FRT0 |
| 34 | 637.8 | 53.5 | 674 | 9 | AI342377 | AI342377 q27608.x |
| 35 | 628.2 | 52.7 | 655 | 10 | BE898853 | BE898853 601682146 |
| 36 | 627.6 | 52.7 | 674 | 13 | B0771685 | B0771685 UI-H-EZ1- |
| 37 | 624.4 | 52.4 | 627 | 12 | BM843483 | BM843483 K-EST0121 |
| 38 | 612.8 | 51.4 | 1187 | 11 | AK002324 | AK002324 Mus muscu |
| 39 | 610.6 | 51.2 | 645 | 14 | CD368900 | CD368900 UI-H-FRT1- |
| 40 | 610.2 | 51.2 | 787 | 10 | BF032665 | BF032665 601453004 |
| 41 | 591 | 49.6 | 627 | 12 | BM919292 | BM919292 AGENCOURT |
| 42 | 580.6 | 48.7 | 1016 | 14 | BY704271 | BY704271 BY704271 |
| 43 | 578 | 48.5 | 638 | 10 | BE613003 | BE613003 601452139 |
| 44 | 574.2 | 48.2 | 688 | 10 | BE271936 | BE271936 601415134 |
| 45 | 570.8 | 47.9 | 1123 | 12 | BI410179 | BI410179 602963868 |

ALIGNMENTS

RESULT 1
LOCUS BM923629 1023 bp mRNA EST 12-MAR-2002
DEFINITION AGENCOURT.6623809 NIH_MGC_116 Homo sapiens CDNA clone IMAGE:5759328
ACCESSION BM923629
VERSION BM923629.1 GI:19374008
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1023)
NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAMA2804 row: j column: 01
High quality sequence stop: 699.

```

FEATURES
  source
    Location/Qualifiers
      1..1023
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:5759328"
        /lab_host="DH10B"
        /clone_lib="NIH_MGC_116"
        /note="Organ: pooled colon, kidney, stomach; Vector:
        PCMV-SORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA
        source anonymous pool of 3 colons, age 26 yo male, 49
        female, 71 yo male colon; 46 yo male kidney, and pool of 2
        stomachs, 62 yo male and 70 yo female. Library is
        oligo-dT primed and directionally cloned (EcoRV site is
        destroyed upon cloning). Average insert size 1.4 kb,
        insert size range 1-3 kb. Library is normalized and
        enriched for full-length clones and was constructed by C.
        Gruber (Invitrogen). Research Genetics tracking code
        023. Note: this is a NIH_MGC Library."
BASE COUNT      192 a      313 c      318 g      196 t      4 others
ORIGIN
Query Match      71.9%; Score 856.6; DB 12; Length 1023;
Best Local Similarity 96.7%; Pred. No. 4.2e-190;
Matches 904; Conservative 0; Mismatches 28; Indels 3; Gaps 3;

19  TCGCTCTCGCAGCCAGCGCGGGGTTTCCGAGAGACAGCCAAAGCATGTCGCCGCG 78
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  5  TTCGATCGCGCAGCCAGCGCGGGGTTTCCAGAGACAGCCAAAGCATGTCGCCGCG 64

79  CCGTTCCGATGTTGGTGGTGGCTACCTCTTGTCTGATTGGCCGCTAGTAGCAA 138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  65  CCGTTCCGATGTTGGTGGTGGCTACCTCTTGTCTGATTGGCCGCTAGTAGCAA 124

139 GATGCTGAGCAAGGCTCTGAAGCGGAACGGGAGAGAGAGAGAGAGAACTCTGCGC 198
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  125 GATGCTGAGCAAGGCTCTGAAGCGGAACGGGAGAGAGAGAGAGAGAACTCTGCGC 184

199 AGTCGACCTCTGTTGGTGGTGGCTACCTCTTGTCTGATTGGCCGCTAGTAGCAA 258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  185 AGTCGACCTCTGTTGGTGGTGGCTACCTCTTGTCTGATTGGCCGCTAGTAGCAA 244

259 GAGCCTCTAGCTCCCTCTTGAACCTCTGAGTGTCAAGCTCCACGACCTGACAGCAAG 318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  245 GAGCCTCTAGCTCCCTCTTGAACCTCTGAGTGTCAAGCTCCACGACCTGACAGCAAG 304

319 TGAGCGGAGCCTGCGGACCTGTTGCTGTTGCTGTAACAATCTGCGGCGCATTCAGGCGTC 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  305 TGAGCGGAGCCTGCGGACCTGTTGCTGTTGCTGTAACAATCTGCGGCGCATTCAGGCGTC 364

379 CATGGCACCCCGCGCTGCTGCCACCTGTGCTAGACCCACCTGACGAGCCGCCCATGTGGCC 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  365 CATGGCACCCCGCGCTGCTGCCACCTGTGCTAGACCCACCTGACGAGCCGCCCATGTGGCC 424

439 TGACAACCTTACCTGGCACTGCGGAGCGTGCCTTTCACCTCCATAGGCGAGCTCTCTCGGA 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  425 TGACAACCTTACCTGGCACTGCGGAGCGTGCCTTTCACCTCCATAGGCGAGCTCTCTCGGA 484

499 GGACCTCAGCCACATTTGAGGGCTGAGTCAAGGCTCCCAACCCCTTGCAGAGAGAGGCGC 558
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  485 GGACCTCAGCCACATTTGAGGGCTGAGTCAAGGCTCCCAACCCCTTGCAGAGAGAGGCGC 544

559 ACCAGGCGCTAGCATGGGGAGAGAGAGCCAGCCCTGGGTGCTTGAACCTGCTGGGCGC 618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  545 ACCAGGCGCTAGCATGGGGAGAGAGAGCCAGCCCTGGGTGCTTGAACCTGCTGGGCGC 604

619 AGCCACTGGCTGTCTACTGAGAGCATGGGCTTGAAGGCTGTGTTAGATATGACACTTC 678
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  605 AGCCACTGGCTGTCTACTGAGAGCATGGGCTTGAAGGCTGTGTTAGATATGACACTTC 664

679 TATGTATGACATGAATCTTTGGGACACAGCTCTGAGGGCTCAAAACAGGAGCCCTGA-GG 737
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  665 TATGTATGACATGAATCTTTGGGACACAGCTCTGAGGGCTCAAAACAGGAGCCCTGAAGG 724

```

```

738 ATGGCCCGGAGAGAGAGAGAGAGTCCGAGTCTGAGAGAGCCGAAATTGGACTACTCATG 797
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  725 ATGGCCCGGAGAGAGAGAGAGAGTCCGAGTCTGAGAGAGCCGAAATTGGACTACTCATG 784

798 ATGTCTGTGTGGCACAACAGCAGCTGAGAGCCAGCCGCGGAGAGAGAGAGAGAGAGAG 856
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  785 ATGTCTGTGTGGCACAACAGCAGCTGAGAGCCAGCCGCGGAGAGAGAGAGAGAGAGAG 844

857 GTGCTGAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 915
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  845 GTGCTGAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 904

916 AAAAGACACAGCTGGCTTCCCTAGTACAGAGACA 950
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  905 AAAAGACACAGCTGGCTTCCCTAATATACCGAANA 939

RESULT 2
B0856974      908 bp      mRNA      linear      EST 16-OCT-2002
LOCUS
DEFINITION
AGENCOURT_10481033 NIH_MGC_107 Homo sapiens cDNA clone
IMAGE:6646774 5', mRNA sequence.
B0856974
ACCESSION
B0856974.1 GI:24041964
VERSION
B0856974.1
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 908)
AUTHORS
NIH-MGC http://mhc.ncl.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.femail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LNCM2886 row: b column: 22
High quality sequence stop: 717.
FEATURES
  source
    Location/Qualifiers
      1..908
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:6646774"
        /tissue_type="adenocarcinoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /clone_lib="NIH_MGC_107"
        /note="Organ: breast; Vector: pOT7; Site:1: EcoRI;
        Site:2: XhoI; cDNA made by oligo-dT priming.
        directionally cloned into EcoRI/XhoI sites using the
        following 5' adaptor: GGCAGAG(G). Library constructed by
        Ling Hong in the laboratory of Gerald M. Rubin (University
        of California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies).
        Note: this is a NIH_MGC Library."
BASE COUNT      182 a      259 c      267 g      170 t      30 others
ORIGIN
Query Match      67.0%; Score 798.8; DB 13; Length 908;
Best Local Similarity 98.4%; Pred. No. 1.4e-176;
Matches 817; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

354 ACACTGTGCGGCGATCCAGGCGTTCATGAGCAGCCGCGGCTGCCCTGACCTGTGCTTA 413
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1  ACACTGTGCGGCGATCCAGGCGTTCATGAGCAGCCGCGGCTGCCCTGACCTGTGCTTA 60

```

| | | | |
|------------|--|--|-----------------|
| QY | 414 | GGCCACCTGCAGGCCGCCAGTGTGGCTGTGACCACTTACTGTGGCAAGCTCCGAGCTGCCCTT | 473 |
| Db | 61 | GCCCACTGCAGGCCGCCAGTGTGGCTGTGACCACTTACTGTGGCAAGCTCCGAGCTGCCCTT | 120 |
| QY | 474 | CAGCCTCATGGCAGCCTCCTGGAGGACCTCAGCCACATTTGAGGGCCTGAGTCAAGCTC | 533 |
| Db | 121 | CAGCCTCATGGCAGCCTCCTGGAGGACCTCAGCCACATTTGAGGGCCTGAGTCAAGCTC | 180 |
| QY | 534 | CCCAACCTTGGCAGAGAGGGGGCCACAGGGCCGTAGATGGGGGGAGCAGCCAGCC | 593 |
| Db | 181 | CCCAACCTTGGCAGAGAGGGGGCCACAGGGCCGTAGATGGGGGGAGCAGCCAGCC | 240 |
| QY | 594 | TGGGTGCTTGGACCTGTGTGGGCCAGGCACTGTGCTGTCTACTGTGAGCATGGGCTTGGAG | 653 |
| Db | 241 | TGGGTGCTTGGACCTGTGTGGGCCAGGCACTGTGCTGTCTACTGTGAGCATGGGCTTGGAG | 300 |
| QY | 654 | GCCGTGTTGAGAGATTTGACACCTCTATGTATGACATGAACTTTGGGACCAGCCTGTG | 713 |
| Db | 301 | GCCGTGTTGAGAGATTTGACACCTCTATGTATGACATGAACTTTGGGACCAGCCTGTG | 360 |
| QY | 714 | AGGCGCTCAAAACAGGCCCTGAGGATGGGGCCAGGGCAAGAGAGAAAGCTCCGAGGTGGACG | 773 |
| Db | 361 | AGGCGCTCAAAACAGGCCCTGAGGATGGGGCCAGGGCAAGAGAGAAAGCTCCGAGGTGGACG | 420 |
| QY | 774 | AGGCGGAATTGACTTACCTCATGGATGTCTGTGTGGGCACACAGGACACTGGAGGACCGC | 833 |
| Db | 421 | AGGCGGAATTGACTTACCTCATGGATGTCTGTGTGGGCACACAGGACACTGGAGGACCGC | 480 |
| QY | 834 | CGGGGCCACAGGGCGGTGAGGCCCTCGTCTGTGGAATGGTTGTCTGAGTATCGAACTAGCCTG | 893 |
| Db | 481 | CGGGGCCACAGGGCGGTGAGGCCCTCGTCTGTGGAATGGTTGTCTGAGTATCGAACTAGCCTG | 540 |
| QY | 894 | CTGGCTGTGACCACTGTCTCTGAAAGACACAGCTGGCTTCCCTAGTACAGAGAAACAGGG | 953 |
| Db | 541 | CTGGCTGTGACCACTGTCTCTGAAAGACACAGCTGGCTTCCCTAGTACAGAGAAACAGGG | 600 |
| QY | 954 | CTTGGGCCACTTTGGAGAGACAGAAATCTAGTCTTGGGCAACTTCATACGCTCTCTCTGT | 1013 |
| Db | 601 | CTTGGGCCACTTTGGAGAGACAGAAATCTAGTCTTGGGCAACTTCATACGCTCTCTCTGT | 660 |
| QY | 1014 | CTCAGGGCTGTGCAGAGGGGAGGCTCGAATTTACCCCTAGTGTGATGGAATGACAGGGCTGTG | 1073 |
| Db | 661 | CTCAGGGCTGTGCAGAGGGGAGGCTCGAATTTACCCCTAGTGTGATGGAATGACAGGGCTGTG | 720 |
| QY | 1074 | GGGACCTGAATTTCCCTGTGGCCCTGTGGGGGTATAT-GCTTGGGCGTTCCTCTCTGATACGGG | 1132 |
| Db | 721 | GGGACCTGAATTTCCCTGTGGCCCTGTGGGGGTATAT-GCTTGGGCGTTCCTCTCTGATACGGG | 780 |
| QY | 1133 | AAGGACCCCCCAATGAGATTTTTCAAATTTAAAGCCAGTCTCGGGAAATCTC | 1182 |
| Db | 781 | AAGGACCCCCCAATGAGATTTTTCAAATTTAAAGCCAGTCTCGGGAAATCTC | 830 |
| RESULT 3 | | | |
| BM558458 | | 1062 bp | linear |
| LOCUS | | | EST 20-PEB-2002 |
| DEFINITION | AGENCOURT_6559395 NIH_MGC_88 Homo sapiens cdna clone IMAGE:5741290 | | |
| ACCESSION | BM558458 | | |
| VERSION | BM558458.1 | | |
| KEYWORDS | EST. | | |
| SOURCE | Homo sapiens (human) | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. | | |
| TITLE | 1 (bases 1 to 1062) | | |
| JOURNAL | NIH-MGC http://mgc.nci.nih.gov/ | | |
| COMMENT | National Institutes of Health, Mammalian Gene Collection (MGC) | | |
| | Unpublished | | |
| | Contact: Robert Strausberg, Ph.D. | | |
| | Email: cgarbs-remail.nih.gov | | |
| | Tissue Procurement: ATCC | | |
| | CDNA Library Preparation: Life Technologies, Inc. | | |

| FEATURES | | | | | | | | | |
|---|-----|--|-----|--|--|--|--|--|--|
| source | | | | | | | | | |
| <p>CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LILAM12757 row: j column: 11 High quality sequence stop: 616 Location/Qualifiers</p> | | | | | | | | | |
| 1. 1062 | | | | | | | | | |
| /organism="Homo sapiens" | | | | | | | | | |
| /mol_type="mRNA" | | | | | | | | | |
| /db_xref="taxon:9606" | | | | | | | | | |
| /clone="IMAGE:5741290" | | | | | | | | | |
| /issue_type="duodenal adenocarcinoma, cell line" | | | | | | | | | |
| /lab_host="DH10B (phage-resistant)" | | | | | | | | | |
| /clone_lib="NIH_MGC_88" | | | | | | | | | |
| /note="Organ: small intestine; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dn primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library." | | | | | | | | | |
| BASE COUNT | | | | | | | | | |
| 204 a 325 c 339 g 190 t 4 others | | | | | | | | | |
| ORIGIN | | | | | | | | | |
| Query Match 66.0%; Score 786.6; DB 12; Length 1062; | | | | | | | | | |
| Best Local Similarity 87.6%; Pred. No. 1.1e-173; | | | | | | | | | |
| Matches 922; Conservative 0; Mismatches 111; Indels 20; Gaps 5 | | | | | | | | | |
| QY | 49 | GAGACAGCCAAACAGCATGCTGCCGCCGCTTCTGATGTTGTTGGGTGGCTAC | 108 | | | | | | |
| Db | 2 | GACACACACCACAAAGCATGCTGCCGCCGCTTCTGATGTTGTTGGGTGGCTAC | 61 | | | | | | |
| QY | 109 | TCTTCGTTCTATTGGCCGTAGTAGCAAGATGCTGAGCAAGGCTCTGAACGGAAACG | 168 | | | | | | |
| Db | 62 | TCTTCGTTCTATTGGCCGTAGTAGCAAGATGCTGAGCAAGGCTCTGAACGGAAACG | 121 | | | | | | |
| QY | 169 | GGAGAGAGAGAGAGAAAGAACTCTGGCAGTGCATCTGTTGGTCTAGATCTGGCCA | 228 | | | | | | |
| Db | 122 | GGAGAGAGAGAGAGAAAGAACTCTGGCAGTGCATCTGTTGGTCTAGATCTGGCCA | 181 | | | | | | |
| QY | 229 | CGCAGCGGTGGCACAGGACCCCGCGCGCTGAGCTCCCTCTTTGACCTCTCAGT | 288 | | | | | | |
| Db | 182 | CGCAGCGGTGGCACAGGACCCCGCGCGCTGAGCTCCCTCTTTGACCTCTCAGT | 241 | | | | | | |
| QY | 289 | GCTCAAGCTCCACCAAGCCTGTCAGCAGAGTAGAGCCGGACCTGCGCACACTGCTGT | 348 | | | | | | |
| Db | 242 | GCTCAAGCTCCACCAAGCCTGTCAGCAGAGTAGAGCCGGACCTGCTGTGTGTGT | 301 | | | | | | |
| QY | 349 | CGTCAACACCTCTGGCGGCCATCTCAAGGCTCCATGGCACCCCGGCTGCCCTGCACCTGT | 408 | | | | | | |
| Db | 302 | CGTCAACACCTCTGGCGGCCATCTCAAGGCTCCATGGCACCCCGGCTGCCCTGCACCTGT | 361 | | | | | | |
| QY | 409 | GCTTAGCCCACTGACAGCCCAAGTGTGGTACACAACTTACGGAAAGCTCGAGACCTGC | 468 | | | | | | |
| Db | 362 | GCTTAGCCCACTGACAGCCCAAGTGTGGTACACAACTTACGGAAAGCTCGAGACCTGC | 421 | | | | | | |
| QY | 469 | CCTTTCAGCCTTCATGCCAGCCTCTCTGAGAGACCTCAAGCCATTTGAGGCTCTAGTCA | 528 | | | | | | |
| Db | 422 | CCTTTCAGCCTTCATGCCAGCCTCTCTGAGAGACCTCAAGCCATTTGAGGCTCTAGTCA | 481 | | | | | | |
| QY | 529 | GGCTCCCAACCCCTTGGGAGAGAGGGGCAACCAAGCGGTACATCGGGGGAGAGAGCGCC | 588 | | | | | | |
| Db | 482 | GGCTCCCAACCCCTTGGGAGAGAGGGGCAACCAAGCGGTACATCGGGGGAGAGAGCGCC | 541 | | | | | | |
| QY | 589 | CAGCCTGGGTGCTTGGACCTGCTGGGCCCAAGCCACTGGCTGTCTACTAGAGAT-GGGC | 647 | | | | | | |
| Db | 542 | CAGCCTGGGTGCTTGGACCTGCTGGGCCCAAGCCACTGGCTGTCTACTAGAGATGGGC | 601 | | | | | | |
| QY | 648 | TTGAGGGCGCTTGGAGATATTGACACCTGTATGATAGACAATAACTTTGGGACAG | 707 | | | | | | |
| Db | 602 | TTGAGGGCGCTTGGAGATATTGACACCTGTATGATAGACAATAACTTTGGGACAG | 661 | | | | | | |
| QY | 708 | CCCTTGAGAGGCTCTAAACCAAGGCCCTTGAGATGAGGCTCGGGCAAGAGAGACGTCGGAGC | 767 | | | | | | |

| | | | |
|----|------|---|------|
| Db | 662 | CCCTGTGAGGGGCTCAAAACCAAGNCCCTGAGGTGGGGCGGCGCAAGAGGAACCTCCGGAGC | 721 |
| QY | 768 | TGGACGAGGGCGGAATTGGACTACTCATCTCATGTGTGTGTGTGTGGGACACACAGGCACCTGGAGC | 827 |
| Db | 722 | TGGACGAGGGCGGAATTGGACTACTCATCTCATGTGTGTGTGTGTGGGACACAGGCACCTGGAGC | 781 |
| QY | 828 | GACCGCCGGGG--CCAGGGCGCTGAGACCTCGCTGCGTGAATGTTGTCTGTG----- | 877 |
| Db | 782 | GACCGCCGGGGCCAGGGCGCTGAGACCTCGCTGTGCTGAATTTGGGTGTCTCTGGNATCTNG | 841 |
| QY | 878 | ATCTGAACGTAGCCTCTGTGCGCTGGAACCAACTGTCTCTGAAAAAGACACAGCTGGCTTCCC | 936 |
| Db | 842 | AACGTAGCGCCCTGCTGGGCGCTGGACCAAACTGTCTCTCCAAAAAGACACAGCTGGCGTTC | 901 |
| QY | 937 | ----TAGTACGAGAAACAGGGGCTTGGGGCACTTTGGAGAGA----CAGAACTCTAGTCTGTG | 988 |
| Db | 902 | CTAGTACAGGAGAAACAGGGGCTTTGGGGCCACTTTGGGAGAGACAGAAATCTATGTCTCG | 961 |
| QY | 989 | GGCAACTTCACATCCGTCTCTCTCTCTCTCAGGGCTGGCAGGGGGAGCCCTGGAATTAACCCC | 1048 |
| Db | 962 | GGGGAACCTTCACATTCCTCCGCCCCCTCGGGGCGCTTAAGGCGCTGGGCAAGGGGGGAACCCC | 1021 |
| QY | 1049 | TAGTGATGGAAATGACAGGGTCTGTGTGGGAACTG | 1081 |
| Db | 1022 | TTTGAATTTTAACCCCTTAACTGATGATGGAATG | 1054 |

| RESULT 4 | LOCUS | DEFINITION |
|----------------|--|-----------------|
| BX101521 | 797 bp, mRNA | linear |
| BX101521 | Soares ovary tumor Nbf01 Homo sapiens cDNA clone | EST 06-FEB-2003 |
| IMAGE998011185 | ; IMAGE:756322, mRNA sequence. | |

| | |
|----------|----------------------|
| SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens |

REFERENCE
AUTHORS
1 (bases 1 to 797)
Ebert, L., Hell, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,

JOURNAL Unpublished
COMMENT Contact: Ina Rolfs

RZPD; IMAGp9980111859.
RZPDLB; I.M.A.G.E. cDNA Clone Collection;

http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl/cgi/responseLibNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .797 |

```

1..797
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE9980111859 ; IMAGE:756322"
/sex="Female"
/issue_type="ovarian tumor"
/lab_host="D110B (ampicillin resistant)"
/clone_id="Scars ovary tumor NDH07"
/node_Organ:ovary:Vector:pRTV3D (Pharmacia) with a
modified polylinker: Site_1: Not I; Site_2: Eco RI; Site_

```

truncated cDNA was primed with a Not I - oligo(dT) primer (5' TGATCCCAATCTCGAAGTCGAGAGCCGCCGCTTTTCTTTTCTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Felina Bonaldo."

[illegible]

LOCUS BU539198 766 bp mRNA linear EST 13-SEP-2002
 DEFINITION AGENCOURT_10215071 NIH_MGC_107 Homo sapiens cDNA clone
 IMAGE:656986 5', mRNA sequence.
 ACCESSION BU539198
 VERSION BU539198.1 GI:22849639
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 766)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHC2757 row: 0 column: 06
 High quality sequence stop: 705.
 Location/Qualifiers
 1. 766
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:656986"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_107"
 /note="Organ: breast; Vector: pOMB7; Site:1; EcoRI;
 Site:2: XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."

FEATURES
 source

BASE COUNT 169 a 209 c 236 g 152 t

Query Match 63.5%; Score 756.4; DB 13; Length 766;
 Best Local Similarity 99.9%; Pred. No. 1.2e-166;
 Matches 757; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

435 TGGCTGACAACTTACTGGCAAGCTCGAGCGTCCCTTCAGCCTCCATGGCCAGCCTCC 494
 1 TGGCTGACAACTTACTGGCAAGCTCGAGCGTCCCTTCAGCCTCCATGGCCAGCCTCC 60
 495 TGGAGACCTCAGCCACATTTGAGGGCTGAGTACGCTCCCAACCTTTGGCAGAGAG 554
 61 TGGAGACCTCAGCCACATTTGAGGGCTGAGTACGCTCCCAACCTTTGGCAGAGAG 120
 555 GGCACACAGCGCTAGCATCGGGGAGACAGCGCCAGCTGGGTGCTTGGACCTGCTGG 614
 121 GGCACACAGCGCTAGCATCGGGGAGACAGCGCCAGCTGGGTGCTTGGACCTGCTGG 180
 615 GGCACACAGCGCTGCTTACTGAGAGATGGGCTTGAGGGCTTGTGAGATATTGACA 674
 181 GGCACACAGCGCTGCTTACTGAGAGATGGGCTTGAGGGCTTGTGAGATATTGACA 240
 675 CCTCTATGATGACATGAACCTTTGGGACACAGCCTCTGAGGGCTCAACACGAGCCTG 734
 241 CCTCTATGATGACATGAACCTTTGGGACACAGCCTCTGAGGGCTCAACACGAGCCTG 300
 735 AGGATGGCGGGGCAAGAGAACTCCGAGCTGAGAGAGAGCGGCAATTGGAATACCTCA 794
 301 AGGATGGCGGGGCAAGAGAACTCCGAGCTGAGAGAGAGCGGCAATTGGAATACCTCA 360

QY 795 TGGATGTGCTGTGGGACACAGGCACTGGAGCGACCCCGGGGCCAGGGCGCTGAGCCC 854
 DB 361 TGGATGTGCTGTGGGACACAGGCACTGGAGCGACCCCGGGGCCAGGGCGCTGAGCCC 420
 QY 855 TGGTGTGGAATGCTGTGCTGATCTGAACTGAGCTGCTGGCTGACCAATGTCCTC 914
 DB 421 TGGTGTGGAATGCTGTGCTGATCTGAACTGAGCTGCTGGCTGACCAATGTCCTC 480
 QY 915 GAAAGACACAGCTGGCTTCCCTAGTACAGAGAAAGAGGCTTGGGCACTTTGGAGAGAC 974
 DB 481 GAAAGACACAGCTGGCTTCCCTAGTACAGAGAAAGAGGCTTGGGCACTTTGGAGAGAC 540
 QY 975 AGAATCTAGTCTGGGCAACTTCACATCCGTCCTCTCTGCTCAGGGCTGGCAGGGAGC 1034
 DB 541 AGAATCTAGTCTGGGCAACTTCACATCCGTCCTCTCTGCTCAGGGCTGGCAGGGAGC 600
 QY 1035 CTGGAAATTACCCCTTGTGATGGAATGACAGGGTCTGTGGAGATCTGATCTCTG 1094
 DB 601 CTGGAAATTACCCCTTGTGATGGAATGACAGGGTCTGTGGAGATCTGATCTCTG 660
 QY 1095 TGGGTCATAGCTTGGGCTGTCTCTCTGATGAGGGAAGAGACCCCAATCAGATTCTT 1154
 DB 661 TGGGTCATAGCTTGGGCTGTCTCTCTGATGAGGGAAGAGACCCCAATCAGATTCTT 720
 QY 1155 CAAATTAAAGCCAGTCTGGGAAATCTCAAAAAAAAAA 1192
 DB 721 CAAATTAAAGCCAGTCTGGGAAATCTCAAAAAAAAAA 758

RESULT 6
 BG715050 803 bp mRNA linear EST 08-MAY-2001
 LOCUS 602673273F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:479569 5',
 DEFINITION mRNA sequence.
 BG715050
 BG715050.1 GI:13993981
 ACCESSION
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 803)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM10679 row: b column: 20
 High quality sequence stop: 802.
 Location/Qualifiers
 1. 803
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4795699"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_96"
 /note="Organ: Brain; Vector: pBluescript (modified
 pBluescript KS+); Site:1: BamHI; Site:2: SalI; XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
 size-selected for average insert size 2.3 kb and
 normalized to R07 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library

FEATURES
 source

constructed by M. Brownstein (NIMH/NRGI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 150 a 249 c 257 g 147 t

ORIGIN

Query Match 63.1%; Score 751.8; DB 10; Length 803; Best Local Similarity 98.7%; Pred. No. 1.4e-165; Matches 789; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

9 GAGTCAGAACTGCTGCTGCGACCCAGGCGGGTTTCCGAGGACAGCCAAACGAT 68
1 GGGTCAGAACTGCTGCTGCGACCCAGGCGGGTTTCCGAGGACAGCCAAACGAT 63
69 GCTGCGCGCGCGCTTCTGATGTTGGTGGGTGGCTTCTTCTGATTTGGCGC 128
64 GCTGCGCGCGCGCTTCTGATGTTGGTGGGTGGCTTCTTCTGATTTGGCGC 123
129 TGGTGAGCAAGATGCTGAGCAAGGCTGAAAGCGGAAAGGAGGAGGAGGAGG 188
124 TGGTGAGCAAGATGCTGAGCAAGGCTGAAAGCGGAAAGGAGGAGGAGGAGG 183
189 AACCTCTGAGCAAGATGCTGAGCAAGGCTGAAAGCGGAAAGGAGGAGGAGG 248
184 AACCTCTGAGCAAGATGCTGAGCAAGGCTGAAAGCGGAAAGGAGGAGGAGG 243
249 CCGCGCGCGCGCTTCTGATGTTGGTGGGTGGCTTCTTCTGATTTGGCGC 308
244 CCGCGCGCGCGCTTCTGATGTTGGTGGGTGGCTTCTTCTGATTTGGCGC 303
309 TGCAGAGAGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 368
304 TGCAGAGAGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363
369 TGCAGAGAGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 428
364 TGCAGAGAGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 423
429 CCAGTGGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 488
424 CCAGTGGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
489 GCGTCTGAGAGAGCTGAGCAAGTGGAGGAGCTGAGTGGAGGAGCTGAGTGGAG 548
484 GCGTCTGAGAGAGCTGAGCAAGTGGAGGAGCTGAGTGGAGGAGCTGAGTGGAG 543
549 ACAGAGGAGCTGAGCAAGTGGAGGAGCTGAGTGGAGGAGCTGAGTGGAGGAG 608
544 ACAGAGGAGCTGAGCAAGTGGAGGAGCTGAGTGGAGGAGCTGAGTGGAGGAG 603
609 TGCAGAGAGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 668
604 TGCAGAGAGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 663
669 TGCAGAGAGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 727
664 TGCAGAGAGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 723
728 GAGCCTGAGAGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 786
724 GAGCCTGAGAGTGGCGGAGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 783
787 CTACTCTGAGAGTGGCTG 805
784 CTA-CTCATGATGTGTG 801

RESULT 7 868 bp mRNA linear EST 24-MAY-2002
LOCUS B0433568
DEFINITION AGNCOURT_7/61470 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020198
ACCESSION B0433568
VERSION B0433568.1 GI:21172644

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE NIH-MGC http://mhc.nhl.nih.gov/.
AUTHORS 1 (bases 1 to 868)
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-riemail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM13223 row: 0 column: 15
High quality sequence stop: 614.
Location/Qualifiers
1. 868
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6020198"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-Sport6; Site: 1; Not; Site: 2; Salt; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.1 kb. Library constructed by Life Technologies."

BASE COUNT 157 a 265 c 284 g 160 t 2 others

ORIGIN

Query Match 62.9%; Score 749.2; DB 13; Length 868; Best Local Similarity 98.7%; Pred. No. 5.9e-165; Matches 818; Conservative 0; Mismatches 4; Indels 7; Gaps 6;

35 GCGCGGGTTTCCGAGAGCAGCCAAAGGATGTCGCGCGCTTCTGATTTGGT 94
1 GCGCGGGTTTCCGAGAGCAGCCAAAGGATGTCGCGCGCTTCTGATTTGGT 60
95 TGTGGTGGCTTCTGATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 154
61 TGTGGTGGCTTCTGATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 120
155 CTGAAGGGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 214
121 CTGAAGGGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
215 CTAGATCTGAGCAG 274
181 CTAGATCTGAGCAG 240
275 TTTGACCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 334
241 TTTGACCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 300
335 CACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 394
301 CACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 360
395 GCGCTGACCTGTCGTCAGCCTGTCAGCCTGTCAGCCTGTCAGCCTGTCAGCCT 454
361 GCGCTGACCTGTCGTCAGCCTGTCAGCCTGTCAGCCTGTCAGCCTGTCAGCCT 420
455 AGCTGAGAGCTGCGCTTTCAGCCTTTCAGCCTTTCAGCCTTTCAGCCTTTCAG 514
421 AGCTGAGAGCTGCGCTTTCAGCCTTTCAGCCTTTCAGCCTTTCAGCCTTTCAG 480
515 GAGGCGCTGAGTCAAGCTCCCAACCTTGGCAGAGAGAGGAGGAGGAGGAGGAG 574

|||||
Db 481 GAGGGCTAGTGGCTCCCAACCCCTTGCGAGAGAGGGGCGACAGCCGTACATC 540
QY 575 GGGGGAGACAGCCCGAGCTGGGTGCTTGAGACCTTGCGGCCAGACCACTGAGTCTA 634
Db 541 GGGGGAGAGCGCCCGACGCTGGGTGCTTGAGACCTTGCGGCCAGACCACTGAGTCTA 600
QY 635 CTGAGAGA-TGGGCTTGAGGGGCTG-TTGGAGATATTTAGACACCTCTATGTATGCAATG 692
Db 601 CTGAGACATTTGGGCTTGAGGGGCTGTTTGGAGATATTTAGACACCTCTATGTATGCAATG 660
QY 693 AACTTTGGGACAGGCTTGAGGGGCTTGAAACAGAG-CCCTGAGATGGCC-GGGCAA 750
Db 661 AACTTTGGGACAGGCTTGAGGGGCTTGAAACAGAGCCCTGAGATGGGCGGGCAA 720
QY 751 GGAGAGAGCTCCGAGAGCTGAGAGAGCCGAATTTGACTACTCATG--ATGTCGTGGTG 808
Db 721 GGAGAGAGCTCCGAGAGCTGAGAGAGCCGAATTTGACTACTCATGAGATGGTGGCTGGTG 780
QY 809 GGCACACAGGCACTGGAGCGACCG-CCGGGGCCAGGGGCGCTGAGCCCTC 856
Db 781 GGCACACAGGCACTGGAGCGACCGCCGGGGCCAGGGGCGCTGAGCCCTC 829

RESULT 8
B0878247 883 bp mRNA linear EST 16-AUG-2002
LOCUS AGENCOURT.8074327 NIH_MGC_110 Homo sapiens CDNA clone IMAGE:6085638
DEFINITION 5', mRNA sequence.
ACCESSION B0878247
VERSION B0878247.1 GI:22270255
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE NIH-MGC http://mhc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cygabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2319 row: f column: 07
High quality sequence stop: 553.
Location/Qualifiers
1. 883

FEATURES
source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6085638"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_110"
/note="Organ: pancreas; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACAG(G). Library constructed by
ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

BASE COUNT 163 a 277 c 285 g 158 t
ORIGIN
Query Match 62.4%; Score 744.2; DB 13; Length 883;
Best Local Similarity 98.0%; Pred. No. 8.7e-164;

Matches 785; Conservative 0; Mismatches 13; Indels 3; Gaps 3;
QY 43 TTTCGAGAGACAGCCAAAGAGATGACCGCCCGCTTCCATTGTTGTTGGG 102
Db 2 TTCTGGAGAGACAGCCAAAGAGATGTCGCCGCCGCTTCTATTGTTGGTGG 61
QY 103 GCTACCTCTTCTGATTGGCCGCTAGTGAACAAGATCTGAGCAAGGGTGAAGCG 162
Db 62 GCTACCTCTTCTGATTGGCCGCTAGTGAACAAGATCTGAGCAAGGGTGAAGCG 121
QY 163 GAAACGGAGAGAGAGAGAGAGAGAACTCTGGCAGTCGACTCTGTTGCTAGATCC 222
Db 122 GAAACGGAGAGAGAGAGAGAGAGAACTCTGGCAGTCGACTCTGTTGCTAGATCC 181
QY 223 TGCCACAGAGCGGTGGCAGAGACCCCGCGCGGCTAGTCCCTCTTACCT 282
Db 182 TGCCACAGAGCGGTGGCAGAGACCCCGCGCGGCTAGTCCCTCTTACCT 241
QY 283 CTGAGTGTCAAGCTTCACACAGCCCTGACAGAGTGAAGCGGACCTGCGACCTGGT 342
Db 242 CTCAGTGTCAAGCTTCACACAGCCCTGACAGAGTGAAGCGGACCTGCGACCTGGT 301
QY 343 GCTGTGTCGAAACACTGTCGCGGCGATCCAGAGGTCCATAGCAACCGCGGCTGCGC 402
Db 302 GCTGTGTCGAAACACTGTCGCGGCGATCCAGAGGTCCATAGCAACCGCGGCTGCGC 361
QY 403 ACCGTGCTTACGCCCACTGACAGCCCGGAGTGGTGAACAATTACTGGCAAGTCCGGA 462
Db 362 ACCGTGCTTACGCCCACTGACAGCCCGGAGTGGTGAACAATTACTGGCAAGTCCGGA 421
QY 463 CGCTGCTTTCAGCTCTCATGAGCCAGCTCTGAGAGACTGACGACATGAGGGCCT 522
Db 422 CGCTGCTTTCAGCTCTCATGAGCCAGCTCTGAGAGACTGACGACATGAGGGCCT 481
QY 523 GAGTCAAGCTTCCCAACCTTGGCAGACAGAGGGGCGCACAGGCCCTGATGGGGAGC 582
Db 482 GAGTCAAGCTTCCCAACCTTGGCAGACAGAGGGGCGCACAGGCCCTGATGGGGAGC 541
QY 583 AGCGCCACAGCTGGGGGCTTGGAGACTGCTGGGGCCAGGACAGTGGTGTCTAGTGAAGA 642
Db 542 AGCGCCACAGCTGGGGGCTTGGAGACTGCTGGGGCCAGGACAGTGGTGTCTAGTGAAGA 601
QY 643 TGGGCTTGAAGGCGCTTGGAGATATTTAGACACCTCTATGTATGACAAATGAATTTGGCC 702
Db 602 TGGGCTTGAAGGCGCTTGGAGATATTTAGACACCTCTATGTATGACAAATGAATTTGGCC 661
QY 703 ACCAGCTTGTGAGGGCTTAAACCAAGCCCTGAGATGGGCGGGAAGAGAGACT-C 761
Db 662 ACCAGCTTGTGAGGGCTTAAACCAAGCCCTGAGATGGGCGGGAAGAGAGACTC 721
QY 762 CGAGAGCTGAGAGA-GGCCGAATTTGACTACTCATGATGTGCTGGGCAACAGAGA 820
Db 722 CGAGAGCTGAGAGAGGCGCAATTTGACTACTCATGATGTGCTGGGCAACAGAGA 781
QY 821 CT-GGAGCGAGCCCGGGGCC 840
Db 782 CTGGGAGCAACCGCGGGGCC 802

RESULT 9
B0537987 765 bp mRNA linear EST 13-SEP-2002
LOCUS AGENCOURT.10186374 NIH_MGC.107 Homo sapiens CDNA clone
DEFINITION IMAGE:6568492 5', mRNA sequence.
ACCESSION B0537987
VERSION B0537987.1 GI:22848428
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NIH-MGC http://mhc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsb@r-mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Plate: L1CM2754 row: e column: 04
High quality sequence stop: 640.
Location/Qualifiers
1. 765
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="6568492"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_107"
/note="Organ: breast; Vector: pOTB7; Site 1: EcoRI; Site 2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(5). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library." 1 others

BASE COUNT 169 a 208 c 235 g 152 t

ORIGIN

Query Match 62.4%; Score 743.4; DB 13; Length 765;
Best Local Similarity 99.6%; Pred. No. 1.3e-163;
Matches 753; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

435 TGGCTGACAACTTACTGCAAGCTGGACGCTGCCCTTTCAGACCTCCATGGCCAGCCTCC 494
1 TGGCTGACAACTTACTGCAAGCTGGACGCTGCCCTTTCAGACCTCCATGGCCAGCCTCC 60
495 TGGAGAGCCTCAGCCACATTGGAGGCTGAGTCAAGCTCCCAACCTTGGCAGAGGAG 554
61 TGGAGAGCCTCAGCCACATTGGAGGCTGAGTCAAGCTCCCAACCTTGGCAGAGGAG 120
555 GGCCACGAGCGCGTAGCATCGGGGAGCAGCGCCAGCCTGGGTGCTTGGACCTGCTGG 614
121 GGCCACGAGCGCGTAGCATCGGGGAGCAGCGCCAGCCTGGGTGCTTGGACCTGCTGG 180
615 GGCCACGAGCGCGTAGCATCGGGGAGCAGCGCCAGCCTGGGTGCTTGGACCTGCTGG 674
181 GGCCACGAGCGCGTAGCATCGGGGAGCAGCGCCAGCCTGGGTGCTTGGACCTGCTGG 240
675 CCTCATGTATGACAAATGAATTTGGGACACAGCCTGAGAGGCTCAAAACAGGCTTG 734
241 CCTCATGTATGACAAATGAATTTGGGACACAGCCTGAGAGGCTCAAAACAGGCTTG 300
735 AGATGAGGCGCGGAGGAGGAGAGCTCGGAGCTGGACGAGCGCAATTGGACTACCTCA 794
301 AGATGAGGCGCGGAGGAGGAGAGCTCGGAGCTGGACGAGCGCAATTGGACTACCTCA 360
795 TGGATGTGCTGGTGGGACACAGGACTGGAGCGACCGCGGGGCGAGGCGCTGAGCCC 854
361 TGGATGTGCTGGTGGGACACAGGACTGGAGCGACCGCGGGGCGAGGCGCTGAGCCC 420
855 TCGAGCTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 914
421 TCGAGCTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
915 GAAAGACACAGCTGGCTCCCTAGTACAGAGAGAGGCTTGGGCCACTTGGAGAGAC 974
481 GAAAGACACAGCTGGCTCCCTAGTACAGAGAGAGGCTTGGGCCACTTGGAGAGAC 540

QY 975 AGAATCTAGTCTGGGCAACCTCACATCCGCTCTCTGCTCAGGGCTGGAGGGGAGC 1034
DB 541 AGAATCTAGTCTGGGCAACCTCACATCCGCTCTCTGCTCAGGGCTGGAGGGGAGC 600
QY 1035 CTGGAATTACCCCTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1094
DB 601 CTGGAATTACCCCTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 660
QY 1095 TGGGCTCAATTCCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1154
DB 661 TGGGCTCAATTCCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 719
QY 1155 CAAATTAAGCCAGCTCTGGGAATTCACAAAAA 1192
DB 720 CAAATTAAGCCAGCTCTGGGAATTCACAAAAA 757

RESULT 10
BG468137 782 bp mRNA linear EST 21-MAR-2001
LOCUS 602509712P1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4644122 5',
DEFINITION mRNA sequence.
ACCESSION BG468137 GI:13400407
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 782)
NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsb@r-mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Plate: L1CM1415 row: g column: 03
High quality sequence stop: 781.
Location/Qualifiers
1. 782
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="4644122"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_15"
/note="Organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(5). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 144 a 243 c 250 g 145 t

ORIGIN

Query Match 62.3%; Score 742.2; DB 10; Length 782;
Best Local Similarity 97.7%; Pred. No. 2.5e-163;
Matches 753; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

28 CGACCCAGGCGCGGCTTCCGAGAGACGCAACAGATGCTGCCGCCGCTTCTCT 87
DB 7 CGACCCAGGCGCGGCTTCCGAGAGACGCAACAGATGCTGCCGCCGCTTCTCT 66
QY 88 GATTGGTTGGGTGGCTTACTCTTCTGATTTGGCCGCTAGTAGAAGATGCTGAG 147

| | | | | |
|------------|--|--------|--|------------------------|
| Db | | 67 | GATTGGTTGTGCGGGAGCTACCTCTTTCGTTTGATTTGGCCGCTAGTGCACAAGATGTGAG | 126 |
| OY | | 148 | CAAAGGCTGTGAAGCGGAAAACGGGAGAGAGAGAGAGAAGAACCTCTGGCAGTCGACTC | 207 |
| Db | | 127 | CAAGGCTCTGAAGCGGAAACGGGAGAGAGAGAGAGAGAAAGAACCTCTGGCAGTCGACTC | 186 |
| OY | | 208 | CTGGTGGCTAGATATCCTGGGCCAAGGAGGGGTGGACACAGGCCACCCCCGGCGGTGGCTTAG | 267 |
| Db | | 187 | CTGGTGGCTAGATATCCTGGGCCAACAAGGGGTGGACACAGGCCACCCCCGGCGGTGGCTTAG | 246 |
| OY | | 268 | CTCCCTCTTTCAGCCTCTCATGTCGTCAAGCTCTCACACAGCCTCTGCAGCAGATGTAGACC | 327 |
| Db | | 247 | CTCCCTCTTTCAGCCTCTCATGTCGTCAAGCTCTCACACAGCCTCTGCAGCAGATGTAGACC | 306 |
| OY | | 328 | CCTGCGGACACCTGGTGTGTCGTGAACACTCTGCGGGGCATCCAGGCGTCCATAGGCACC | 387 |
| Db | | 307 | CCTGCGGACACCTGGTGTGTCGTGAACACTCTGCGGGGCATCCAGGCGTCCATAGGCACC | 366 |
| OY | | 388 | CGCGGCTGCCCTGGCCACTCTGTGCTTAGCCCACTGTGACGCCCCAGGTGTGGCTGACAACTT | 447 |
| Db | | 367 | CGCGGCTGCCCTGGCCACTCTGTGCTTAGCCCACTGTGACGCCCCAGGTGTGGCTGACAACTT | 426 |
| OY | | 448 | ACTGGCAAAGCTCGGAGCGTGGCCCTTTCAGCCTCTCATGAGCCGACGCTCTGGAGAGCCTCAG | 507 |
| Db | | 427 | ACTGGCAAAGCTCGGAGCGTGGCCCTTTCAGCCTCTCATGAGCCGACGCTCTGGAGAGCCTCAG | 486 |
| OY | | 508 | CCACATTTAGAGGGCTGAGTCAAGGCTCCCAACCCCTTGGCAGACGAGGGGCCACAGGCGCG | 567 |
| Db | | 487 | CCACATTTAGAGGGCTGAGTCAAGGCTCCCAACCCCTTGGCAGACGAGGGGCCACAGGCGCG | 546 |
| OY | | 568 | TAGCATGAGGGGGAGACAGCGCCACGCTGTGGTGTGGACTCTGCTGGGCCACAGCCACTGG | 627 |
| Db | | 547 | TAGCATGAGGGGGAGACAGCGCCACGCTGTGGTGTGGACTCTGCTGGGCCACAGCCACTGG | 606 |
| OY | | 628 | CTGCTCTACTGAGCAGTGGGCTTTGAGGGGCGCTTTTGAGGATTTTGACACTCTATGTATAGA | 687 |
| Db | | 607 | CTGCTCTACTGAGCAGTGGGCTTTGAGGGGCGCTTTTGAGGATTTTGACACTCTATGTATAGA | 666 |
| OY | | 688 | CAATGAACCTTTGGGACACAGCCCTCTGAGGGGCTCAAAACAGGCCCCTGAGATGGGCGGG | 747 |
| Db | | 667 | CAATGAACCTTTGGGACACAGCCCTCTGAGGGGCTCAAAACAGTGGCTGAGATGGGCGGG | 726 |
| OY | | 748 | CAAGGAGGAACTCCGAGAGCTGGACAGAGGCCGAATTGGACTTAACCTATGGA | 798 |
| Db | | 727 | CAGAGGAAAGCTCCGAGACTTGGACAGAGGTGAATTGGACTTAACCTATGGA | 777 |
| RESULT_11 | | | | |
| BGB829426 | | | | |
| LOCUS | BGB829426 | 846 bp | mRNA | linear EST 22-MAY-2001 |
| DEFINITION | 6027637J1F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4899146 5', | | | |
| VERSION | BGB829426 | | | |
| KEYWORDS | BGB829426.1 GI:14177013 | | | |
| SOURCE | EST. | | | |
| ORGANISM | Homo sapiens (human) | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | |
| TITLE | NIH-MGC http://mgc.nci.nih.gov/. | | | |
| JOURNAL | National Institutes of Health, Mammalian Gene Collection (MGC) | | | |
| COMMENT | Unpublished Contact: Robert Strausberg, Ph.D. Email: cgabbs@femail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov | | | |

| | Plate: L16M1790 | row: 1 | column: 03 |
|-----------------------|---|--------|------------|
| | High quality sequence stop: 820. | | |
| FEATURES | Location/Qualifiers | | |
| Source | 1. .846 /organism="Homo sapiens" /mol_type="mrna" /db_xref="taxon:9606" /clone="IMAGE:4899146" /tissue_type="epithelioid carcinoma cell line" /lab_host="DH10B (phage-resistant)" /clone_1lb="NIH_MGC_42" /note="Organ: pancreas; Vector: pCMV7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC library." | | |
| BASE COUNT | 159 a 260 c 276 g 151 t | | |
| ORIGIN | | | |
| Query Match | 62.1%; Score 740.8; DB 12; Length 846; | | |
| Best Local Similarity | 95.8%; Pred. No. 5.4e-163; | | |
| Matches | 805; Conservative 0; Mismatches 27; Indels 8; Gaps 4; | | |
| Dn | 27 GCACGCCAGCGCGGGTTCCGGAGAGCAGCCAAACAGCATGTCTGCCGCCGTTTTC 86 | | |
| Dn | 2 GCGACCACGGGGCGGTTCCGGAGAGCAGCCAAACAGCATGTCTGCCGCCGTTTTC 61 | | |
| Oy | 87 TGATTGGTTGGTGCGCTACCTCTTCGTTGATTTGGCCGCTAGTAGCAATGCTGA 146 | | |
| Dn | 62 TGATTGGTTGGTGCGCTACCTCTTCGTTGATTTGGCCGCTAGTAGCAATGCTGA 121 | | |
| Oy | 147 GCAAGGGCTGTAACGGGAAGGGAGAGAGAGAGAAGGAACCTCTGGCAGTCACT 206 | | |
| Dn | 122 GCAAGGGCTGTAACGGGAAGGGAGAGAGAGAAGGAACCTCTGGCAGTCACT 181 | | |
| Oy | 207 CCTGGTGGCTAGATCTTGCGCCACGACGAGGGTGGACAAGGCAACCCCGCCCTGGGCTCTA 266 | | |
| Dn | 182 CCGTGGCTAGATCTTGCGCCACGACGAGGGTGGACAAGGCAACCCCGCCCTGGGCTCTA 241 | | |
| Oy | 267 GCTCCCTCTTTGACCTCTCACTGCTCAAAGCTCACACACGCTCGACAGAGTAGACGG 326 | | |
| Dn | 242 GCTCCCTCTTTGACCTCTCACTGCTCAAAGCTCACACACGCTCGACAGAGTAGACGG 301 | | |
| Oy | 327 ACCTGGCGGACCTGGTGTGTCGTGCGTAACACTCTGCGCGGATTCAGGCGTTCATGGAC 386 | | |
| Dn | 302 ACCTGGCGGACCTGGTGTGTCGTGCGTAACACTCTGCGCGGATTCAGGCGTTCATGGAC 361 | | |
| Oy | 387 CCGGGGGCGCCCTGGCACCTGTGCTTACGCCACCCAGCCGACGCCAGTGTGGCTGCACT 446 | | |
| Dn | 362 CCGGGGGCGCCCTGGCACCTGTGCTTACGCCACCCAGCCGACGCCAGTGTGGCTGCACT 421 | | |
| Oy | 447 TACTGGCAAGCTGGAGCGCTGCCCTTTCAGCTCCATGCGCAGGCTCTGGAGAGACTCA 506 | | |
| Dn | 422 TACTGGCAAGCTGGAGCGCTGCCCTTTCAGCTCCATGCGCAGGCTCTGGAGAGACTCA 481 | | |
| Oy | 507 GCCACATTGAGGGGCTGATGAGGCTCCCCAACCCCTTGGCAGAGAGGGGCCACCAAGCC 566 | | |
| Dn | 482 GCCACATTGAGGGGCTGATGAGGCTCCCCAACCCCTTGGCAGAGAGGGGCCACCAAGCC 541 | | |
| Oy | 567 GTACACATGGGGGAGACAGCGCCAGCCAGCTGGGATGCTGGACCTGTGGGCCACGCACTG 626 | | |
| Dn | 542 GTACACATGGGGGAGACAGCGCCAGCCAGCTGGGATGCTGGGACCTGTGGGCCACGCACTG 601 | | |
| Oy | 627 GCTGCTACTGGACGATGGGCTTGAAGGCGCTGTTGAGATATTGACACCTCTATGATG 686 | | |
| Dn | 602 GCTGCTACTGGACGATGGGCTTGAAGGCGCTGTTGAGATATTGACACCTCTATGATG 661 | | |
| Oy | 687 ACAATGAACCTTTGGGACACAGCCTTCGTGAGGGGCTTC-AAACGAGGCCCTGAGGATGGCGG 745 | | |

Db 662 ACAATGAACCTTGGGACACAGCCTCTGTAGGGCCCTCAAAAACAGGCCCTGTAGANGATGGCCG 721

QY 746 GG--CAAGCAGAAAGCTCCGAGAGTGTAGACGAGGCCGAATTGGACTTACCTCATGTATGTG- 802

Db 722 GGACAGAGAGAAAGCTCCGGAGCTGACGAGGTGTGAATTGGACTTACCTCATGTATGTG- 781

QY 803 --CTGGTGGGACACAGGACT--GGAGCAGACCCGGGGCCAGGGCCGTGAGCCCTGT 858

Db 782 GCGTGTGGGACACAGGACTTGGAGCGAACCGGGGGCCAGAGGCCCTGAGACATCGT 841

RESULT 12

LOCUS CD364905/c 747 bp mRNA linear EST 29-MAY-2003

DEFINITION UI-H-FT2-bjn-c-13-0-UI.s1 NCI_CGAP_Ft2 Homo sapiens cDNA clone

ACCESSION CD364905

VERSION CD364905.1 GI:31148995

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 747)

TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Gary W. Hunninghake, U of I

CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at <http://genome.uiowa.edu/distribution/cgap.html>

Seq primer: M13 FORWARD

POLYA=yes

FEATURES

source

Location/Qualifiers

1..747

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-H-FT2-bjn-c-13-0-UI"

/tissue_type="Aveolar Macrophage"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies)"

/clone_id="NCI_CGAP_Ft2"

/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker. Site_1: EcoR I; Site_2: Not I. NCI_CGAP_Ft2 is a subcloned cDNA library constructed from a pool of 81 RNA samples from Alveolar Macrophages challenged with different treatments. The library was subtracted according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. The tissue was provided by Dr. Gary W. Hunninghake of the University of Iowa.

TAG_LIB=UI-H-FT2

TAG_TISSUE=Human Lung Aveolar Macrophage

TAG_SEQ=GGCCAGCCG

BASE COUNT 146 a 229 c 204 g 166 t 2 others

ORIGIN

Query Match 61.88; Score 737; DB 14; Length 747;

Best Local Similarity 99.78; Pred. No. 4.1e-162;

Matches 737; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 454 AAGCTCGAGCGCTCCCTTTCAGGCTCATGAGCCAGGCTCTGTAGAGACCTCAGCCACAT 513

Db 747 AAGCTCGAGCGCTCCCTTTCAGGCTCATGAGCCAGGCTCTGTAGAGACCTCAGCCACAN 688

QY 514 TGAGGGCTGAGTGAAGCTCCCAACCTTGGCAGAGAGGGCCACAGGCGCTAGCAT 573

Db 687 TGAGGGCTGAGTGAAGCTCCCAACCTTGGCAGAGAGGGCCACAGGCGCTAGCAT 628

QY 574 CGGGGAGCAGCGCCACAGCTTGCTGCTGGACCTGCTGGGCCACAGCCACTGGCTGCT 633

Db 627 CGGGGAGCAGCGCCACAGCTTGCTGCTGGACCTGCTGGGCCACAGCCACTGGCTGCT 568

QY 634 ACTGAGCATGGGCTTGAGGGCTGTTTGGAGATTTGACACCTCTATGTATGACATGA 693

Db 567 ACTGAGCATGGGCTTGAGGGCTGTTTGGAGATTTGACACCTCTATGTATGACATGA 508

QY 694 ACTTGGGACAGCGCTCTGAGGGGCTCAAAACAGGGCCCTGAGATGGGCGGCGCAAGA 753

Db 507 ACTTGGGACAGCGCTCTGAGGGGCTCAAAACAGGGCCCTGAGATGGGCGGCGCAAGA 448

QY 754 GGAAGCTCCGAGCTGAGCAGAGGCCGAATTGACTTACCTCAATGATGCTGTGGGCAC 813

Db 447 GGAAGCTCCGAGCTGAGCAGAGGCCGAATTGACTTACCTCAATGATGCTGTGGGCAC 388

QY 814 ACAGGACATGAGAGGACCGCCGGGGCCAGGGCGCTGAGCCCTGCTGTGAAATGGTGTG 873

Db 387 ACAGGACATGAGAGGACCGCCGGGGCCAGGGCGCTGAGCCCTGCTGTGAAATGGTGTG 328

QY 874 TGGTATCTGAACTGAGCCTGCTGGCTGGACCACTGCTCGAAAGACAGCGTGGCT 933

Db 327 TGGTATCTGAACTGAGCCTGCTGGCTGGACCACTGCTCGAAAGACAGCGTGGCT 268

QY 934 CCTAGTACAGAGAAACAGGCTTGGGCCACTTGGAGAGACAGATCTAGTCTGGGCAA 993

Db 267 CCTAGTACAGAGAAACAGGCTTGGGCCACTTGGAGAGACAGATCTAGTCTGGGCAA 208

QY 994 CTTACACATCCCTCTCTGCTGCTGCTGAGGCTGGCAGAGGGAGAGGCTGGAATACCCCTAGTG 1053

Db 207 CTTACACATCCCTCTCTGCTGCTGCTGAGGCTGGCAGAGGGAGAGGCTGGAATACCCCTAGTG 148

QY 1054 ATGGAATGACAGGCTGCTGGGAGAGTGAATTCCTGCGCTGGGGCTATGAGTGGGCT 1113

Db 147 ATGGAATGACAGGCTGCTGGGAGAGTGAATTCCTGCGCTGGGGCTATGAGTGGGCT 88

QY 1114 GTTCCTCTCTGATACGGGAAAGAGACCCCAATCAGATTTTCAATTAAAGCCAGTCTG 1173

Db 87 GTTCCTCTCTGATACGGGAAAGAGACCCCAATCAGATTTTCAATTAAAGCCAGTCTG 28

QY 1174 GGAATCTCAAAAAA 1192

Db 27 GGAATCTCAAAAAA 9

RESULT 13

LOCUS BG402912 1037 bp mRNA linear EST 12-MAR-2001

DEFINITION 602418715F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4525995 5', mRNA sequence.

ACCESSION BG402912

VERSION BG402912.1 GI:13296360

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 1037)

TITLE NIH-MGC <http://mhc.ncbi.nlm.nih.gov/>.

COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

CONTACT: Robert Strausberg, Ph.D.

COMMENT Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

plate: LLNL0432 row: m column: 04

High quality sequence stop: 739.

[illegible]

| FEATURES | source | Location/Qualifiers |
|-----------------------|-------------------|---|
| 1. | 949 | /organism="Homo sapiens" |
| | | /mol_type="mRNA" |
| | | /db_xref="taxon:9606" |
| | | /clone="IMAGE:6731456" |
| | | /tissue.type="adenocarcinoma, cell_line" |
| | | /lab_host="DH10B (phage-resistant)" |
| | | /clone_idb="NIH_MGC_107" |
| | | /note="Organ: breast; Vector: pORF7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library." |
| BASE COUNT | 221 a | 264 c 289 g 174 t 1 others |
| ORIGIN | | |
| Query Match | 61.0%; | Score 727; DB 13; Length 949; |
| Best Local Similarity | 94.7%; | Pred. No. 9,5e-160; |
| Matches | 809; Conservative | 0; Mismatches 31; Indels 14; Gaps 5; |
| QY | 353 | AACACTGCGCGCGCATCCAGGCGCTGCATATGGCACCCGCGGCTGCCTCCACACTGTGCT 412 |
| Db | 2 | AACACTGCGCGCGCATCCAGGCGCTGCATATGGCACCCGCGGCTGCCTCCACACTGTGCT 61 |
| QY | 413 | AGCCCACTCGAGGCCCCAGTGTGGCTATACAATCTACTGCGAAGTGGAGAGCTGCTT 472 |
| Db | 62 | AGCCCACTCGAGGCCCCAGTGTGGCTATACAATCTACTGCGAAGTGGAGAGCTGCTT 121 |
| QY | 473 | TCAGCCTTCAGAGGCGCAGCTCTGTGGAGAGCTCAGCCACATGTGAGAGGCGCTAGTCAGGCT 532 |
| Db | 122 | TCAGCCTTCAGAGGCGCAGCTCTGTGGAGAGCTCAGCCACATGTGAGAGGCGCTAGTCAGGCT 181 |
| QY | 533 | CCCCAACCCCTTGGCAGACGAGGGGGCCACAGGCGCTAGCATCGGGGGAGCAAGCGCCAGC 592 |
| Db | 182 | CCCCAACCCCTTGGCAGACGAGGGGGCCACAGGCGCTAGCATCGGGGGAGCAAGCGCCAGC 241 |
| QY | 593 | CTGGGTGCTTGGAGCACTCTGGGGCCAGCCACTGGCTGTCTACTGAGAGAGTGGGCTTAG 652 |
| Db | 242 | CTGGGTGCTTGGAGCACTCTGGGGCCAGCCACTGGCTGTCTACTGAGAGAGTGGGCTTAG 301 |
| QY | 653 | GGCCTGTTTGAAGATTTGACACCTCTATGTATGACATGAACTTTGGGCAACAGCTCT 712 |
| Db | 302 | GGCCTGTTTGAAGATTTGACACCTCTATGTATGACATGAACTTTGGGCAACAGCTCT 361 |
| QY | 713 | GAGGGCTCAAAACAGGCGCTGAGAGATGGCCGGGCAAGAGGAACTCCGGAGCTGGAC 772 |
| Db | 362 | GAGGGCTCAAAACAGGCGCTGAGAGATGGCCGGGCAAGAGGAACTCCGGAGCTGGAC 421 |
| QY | 773 | GAGGCGCAATTTGAGCTACTCATGTGATGGTGTGGGCAACAGGCACTGGAGCGAGCG 832 |
| Db | 422 | GAGGCGCAATTTGAGCTACTCATGTGATGGTGTGGGCAACAGGCACTGGAGCGAGCG 481 |
| QY | 833 | CGGGGGCCAGGGCGCTGAGCCCTGCTCGTGGAAATGGTTGTCTGTATCTGAATGAGCT 892 |
| Db | 482 | CGGGGGCCAGGGCGCTGAGCCCTGCTCGTGGAAATGGTTGTCTGTATCTGAATGAGCT 541 |
| QY | 893 | GGTGGCGAGCAACTGCTCCGAAAGAGACACACTGGCTCCCTACTACAGAGAAACAGG 952 |
| Db | 542 | GGTGGCGAGCAACTGCTCCGAAAGAGACACACTGGCTCCCTACTACAGAGAAACAGG 601 |
| QY | 953 | GCTTGGGCGCACTTTGGAGAGCAANAATCTAGTCTGGGCAACTTCACATCGTCTCTCG 1012 |
| Db | 602 | GCTTGGGCGCACTTTGGAGAGCAANAATCTAGTCTGGGCAACTTCACATCGTCTCTCG 661 |
| QY | 1013 | TCTTCAGGGGCTGGCAGGGGGAGCCTTGAATTAACCCCTAGTGTGATGAAATGACAGGG - TCT 1070 |
| Db | 662 | TCTTCAGGGGCTGGCAGGGGGAGCCTTGAATTAACCCCTAGTGTGATGAAATGACAGGGGCTG 721 |


```
QY      1071 GGTGGGACTGAAATCCCTGGCCC--TGGGTCATAGCTTGGGCTGTCTCTCT---G 1125
Db      722  GTGGGGGACTGAATTTCTGNCCTCGGGGCTCATACCTTGGGCTGTCTCTCTGA 781
QY      1126 ATACGGGAAGAGACCCCA---ATCAGATTTTCAAAATTAAGCC---AGTCCTGGGAAA 1178
Db      782  AAGGGGAAGAGACCCCAATCAGAAATTTTCAATTAAGCCCAAGTCTCTGGGGAAA 841
QY      1179 TCTCAAAAAAAAAA 1192
Db      842  TCTCAAAAAAAAAA 855
```

Search completed: August 25, 2003, 16:45:27
job time : 2467 secs

THIS PAGE BLANK (USPTO)

Db 361 GCCGGCTGGACCCAGGCTCCAGGCGCGCGCGCGGAGAGACGAAACGCCGAG 302
QY 512 ATTGAGGCGCTGAGTACGCTCCCAACCTTGGACAGAGGCGGCACAGCCGCTAGG 571
Db 301 GGGGCGCGGACAGAGGCGGCGGCTCCAGGCTGCTCCAGGCGCGGACGACGCTAGC 242
QY 572 ATCGGGGAGACGCGCCGAGCC 593
Db 241 GCTGGCTGGCGCGCGGAGGCG 220

RESULT 5
US-09-252-991A-8324
; Sequence 8324, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8324
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8324

Query Match 3.7%; Score 43.6; DB 4; Length 1278;
Best Local Similarity 43.7%; Pred. No. 0.069;
Matches 193; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

QY 152 GGTCTGAGCGGAAACGGAGAGAGAGAGAAACCTCTGGCAGTGCACCTCTGG 211
Db 168 GGTCTGAGCGGCGCTCCGCGGAGGCTCCAGAGGCGGCGCGCGCTCCGCGGAG 227
QY 212 TGGCTGATCTTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 271
Db 228 CAGCAGGCTGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 287
QY 272 CTTTGAACCTTCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTG 331
Db 288 GCGGCGGCTCTCTTCACTTCCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 347
QY 332 CCGACCTGAGTGTGCTGTAACACTTGGCGGAGGAGGAGGAGGAGGAGGAGGAG 391
Db 348 GCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 407
QY 392 GCTGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 451
Db 408 GCGCAGCAACCAATGCGGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 467
QY 452 GCAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 511
Db 468 GCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 527
QY 512 ATTGAGGCGCTGAGTACGCTCCCAACCTTGGACAGAGGCGGCACAGCCGCTAGC 571
Db 528 GCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 587
QY 572 ATCGGGGAGAGGCGCCGAGCC 593
Db 588 GCTGGCTGGCGCGCGGAGGCG 609

RESULT 6
US-09-252-991A-8483/c

; Sequence 8483, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8483
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8483

Query Match 3.7%; Score 43.6; DB 4; Length 1362;
Best Local Similarity 43.7%; Pred. No. 0.071;
Matches 193; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

QY 152 GGTCTGAGCGGAAACGGAGAGAGAGAGAGAAACCTCTGGCAGTGCACCTCTGG 211
Db 1029 GGTCTGAGCGGCGCTCCGCGGAGGCTCCAGAGGCGGAGGAGGAGGAGGAGGAG 970
QY 212 TGGCTGATCTTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 271
Db 969 CAGCAGGCTGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 910
QY 272 CTTTGAACCTTCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTG 331
Db 909 GCGGCGGCTCTCTTCACTTCCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 850
QY 332 CCGACCTGAGTGTGCTGTAACACTTGGCGGAGGAGGAGGAGGAGGAGGAGGAG 391
Db 849 GCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 790
QY 392 GCTGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 451
Db 789 GCGCAGCAACCAATGCGGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 730
QY 452 GCAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 511
Db 729 GCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 670
QY 512 ATTGAGGCGCTGAGTACGCTCCCAACCTTGGACAGAGGCGGCACAGCCGCTAGC 571
Db 669 GCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 610
QY 572 ATCGGGGAGAGGCGCCGAGCC 593
Db 609 GCTGGCTGGCGCGCGGAGGCG 588

RESULT 7
US-09-252-991A-8356
; Sequence 8356, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142


```

: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 14680
: LENGTH: 1095
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14680

```

| | | | | |
|---------------------------|-------|----------------|----------|-------------|
| Query Match | 3.3% | Score 38.8 | DB 4 | Length 1095 |
| Best Local Similarity | 47.2% | Pred. No. 1.1 | | |
| Matches 118; Conservative | 0 | Mismatches 132 | Indels 0 | Gaps 0 |

| | | | |
|----|-----|---|-----|
| QY | 151 | GGGCTCGAAGCGGAAACGGGAGAGAGAGAGAAGAACCTCTGGCAGTCGACTCTG | 210 |
| | 425 | GGCCCTCGAAGCAGAGAACTACGTGCTAGTGTTCGAAATGCGGACCGACGCCGCCCGGGAG | 366 |
| Db | 211 | GTGGCTAGATCTCTGGCCACGACGCGGTGGACAGGCAACCCCGGCCCTTGCTTACCTC | 270 |
| | 365 | CGCGCTGATGTGATCATCATGCTGTGATTCGCGACACTGGGACGGAAAGGGAAATTGTCGA | 306 |
| Db | 271 | CTCTTTTGACCTCTCAGTGTGCTCACTCCACACAGCCTGGACGAGAGATGAGCCGGACCT | 330 |
| QY | 305 | GCAGTTGGGGGACTTCGGGAGCAACTGGTGTACCGCCTTGACACAGCAGGGTTCAGTCT | 246 |
| Db | 331 | GGGGCACTGGTGTGCTGTGTAACACTCTTGCGGGCATTCGACGTCATGCACTGCACCCGC | 390 |
| QY | 245 | GTTCCTGGGAGGTCGTGATCTCTCTCTGTGAACGGGCGCAGGCGCATTTCCACAGGGG | 186 |
| Db | 391 | GGCTGGCCCTG | 400 |
| QY | 185 | TTTTTCCCGTG | 176 |

```

RESULT 11
US-09-252-991A-14748
Sequence 14748, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196, 136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 14748
LENGTH: 1317
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14748

```

| | | | | |
|-----------------------|----------------|----------------|----------|-------------|
| Query Match | 3.3% | Score 38.8 | DB 4 | Length 1317 |
| Best Local Similarity | 47.2% | Pred. No. 1,2 | | |
| Matches 118 | Conservative 0 | Mismatches 132 | Indels 0 | Gaps 0 |

| | | | |
|----|-----|--|-----|
| QY | 151 | GGGCTCAACGGGAAACGGAGAGAGAGAGAGAACCTCTGGCAGTGCACCTCTG | 210 |
| | | | |
| Db | 876 | GGCCCTCAAGAGAGAAATACGTCGTAGTGTGTGAATGTGGACCCAGCGCCCGCGGGAG | 935 |
| | | | |
| QY | 211 | GTGGCTAGATCTTGGCCACGACGCGGTGGACAGGACCCCGCGCGCTTACGCTC | 270 |

Db 936 CGGGGTAGCTATATATCACTGCTGCATGCCCCAGCTGGGAGGAAAGCGGAATTGCTTCA 995

QY 271 CCTCTTTGACCTCTCAAGTGTCAAGCTCCACAGCCTGGAGCAGATGACGGGACCT 330

Db 996 GCAGTTGGGGCACTTCCGGGAGCAAGTGGTTACCGGCTGCACCAAGCAGGGTCCAGTTCCT 1055

QY 331 GCGGCACCTGGTGGTGTGTGGAACACTCTGCGGATCCAGAGCGTTCATGGACCCGC 390

Db 1056 GTTCCGGCAGGTGCTGGTAGTCTCTCTCTGGAACGGGACCAGGGGCATTTCACACGGGG 1115

QY 391 GGCTGCCCTTG 400

Db 1116 TTTTCCCGTG 1125

```

RESULT 12
US-09-501-115-31
; Sequence 31, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Fader, Gary M.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Chlnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: BB1338 US NA
; CURRENT APPLICATION NUMBER: US/09/501.115
; CURRENT FILING DATE: 2000-02-09
; EARLIER APPLICATION NUMBER: 60/119,585
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 31
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Triticum aestivum
; US-09-501-115-31

```

[illegible]

```

RESULT 13
US-09-370-838-153
; Sequence 153, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roachd
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE OF INVENTION: LONG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370. 838
; CURRENT FILING DATE: 1999-08-09

```


.,Tue Aug 26 13:49:14 2003

us-09-701-675a-8.rni

Page 7

Search completed: August 25, 2003, 16:47:17
Job time : 104 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 12:16:56 ; Search time 371 Seconds
(without alignments)
8673.127 Million cell updates/sec

Title: US-09-701-675a-8

Perfect score: 1192
Sequence: 1 gtagcgagcgagtcagactg.....gggaatctcaaaaaaaaa 1192

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

N_Geneseq_19jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 1192 | 100.0 | 1192 | 21 | AAZ29482 |
| 2 | 1188.8 | 99.7 | 1219 | 22 | AAH64895 |
| 3 | 1183.4 | 99.3 | 2113 | 22 | AAK94674 |
| C | 1167 | 97.9 | 1181 | 22 | AAH99491 |
| C | 1154.4 | 96.8 | 1222 | 21 | AACT6627 |
| 5 | 1151.2 | 96.6 | 1182 | 20 | AAK97688 |
| 7 | 1061.2 | 89.0 | 1078 | 20 | AAZ19894 |
| C | 889.4 | 74.6 | 1216 | 23 | AAZ78542 |

| | | | | | | |
|---|----|-------|------|--------|----|----------|
| C | 9 | 777.8 | 65.3 | 5746 | 24 | ABK81819 |
| C | 10 | 707.8 | 59.4 | 711 | 20 | AAK36177 |
| C | 11 | 667.4 | 56.0 | 5869 | 24 | AAI49936 |
| C | 12 | 637.2 | 53.5 | 696 | 22 | AAK91707 |
| C | 13 | 637.2 | 53.5 | 696 | 22 | AAK93731 |
| C | 14 | 447 | 37.5 | 447 | 21 | AAZ29502 |
| C | 15 | 433.8 | 36.4 | 450 | 20 | AAV86113 |
| C | 16 | 293.4 | 24.6 | 301 | 20 | AAV41286 |
| C | 17 | 237.6 | 19.9 | 336 | 24 | AAZ62144 |
| C | 18 | 232 | 19.5 | 363 | 24 | AAZ11398 |
| C | 19 | 189 | 15.9 | 189 | 21 | AAZ29488 |
| C | 20 | 188 | 15.8 | 190 | 21 | AAZ29504 |
| C | 21 | 120.8 | 10.1 | 165 | 24 | ABZ69457 |
| C | 22 | 60 | 5.0 | 60 | 24 | ABK39710 |
| C | 23 | 52.2 | 4.4 | 65 | 24 | ABK55795 |
| C | 24 | 49.2 | 4.1 | 1253 | 24 | ABK97063 |
| C | 25 | 44.6 | 3.7 | 862 | 22 | AAH06538 |
| C | 26 | 44.6 | 3.7 | 1314 | 24 | ABL49854 |
| C | 27 | 44.6 | 3.7 | 2082 | 24 | AAI41498 |
| C | 28 | 44.6 | 3.7 | 2447 | 24 | AAZ34182 |
| C | 29 | 44.6 | 3.7 | 2629 | 22 | AAI58881 |
| C | 30 | 44.6 | 3.7 | 2629 | 22 | AAI60667 |
| C | 31 | 44.2 | 3.7 | 4356 | 14 | AAQ37543 |
| C | 32 | 43 | 3.6 | 396 | 22 | AAE94876 |
| C | 33 | 43 | 3.6 | 396 | 24 | ABT03243 |
| C | 34 | 43 | 3.6 | 396 | 24 | ABL48926 |
| C | 35 | 43 | 3.6 | 2038 | 21 | AAE18134 |
| C | 36 | 43 | 3.6 | 2038 | 22 | AAH33113 |
| C | 37 | 42.6 | 3.6 | 4356 | 16 | AAQ95540 |
| C | 38 | 42.4 | 3.6 | 1387 | 13 | AAQ20688 |
| C | 39 | 42.4 | 3.6 | 1387 | 21 | AAZ49675 |
| C | 40 | 42.2 | 3.5 | 1428 | 25 | AAZ53024 |
| C | 41 | 42.2 | 3.5 | 2082 | 24 | AAI41497 |
| C | 42 | 42.2 | 3.5 | 122748 | 24 | ABT10719 |
| C | 43 | 41.8 | 3.5 | 1413 | 25 | AAZ53019 |
| C | 44 | 41.4 | 3.5 | 2949 | 25 | ABZ71137 |
| C | 45 | 41.4 | 3.5 | 36321 | 25 | ABZ71131 |

ALIGNMENTS

| | | |
|----------|---|---|
| RESULT 1 | AAZ29482 | standard; DNA; 1192 BP. |
| ID | AAZ29482 | |
| XX | AAZ29482; | |
| AC | AAZ29482; | |
| XX | | |
| DT | 14-MAR-2000 | (first entry) |
| DE | DNA encoding human cell cycle regulation protein-3. | |
| XX | | |
| KW | CEBP-3; cell cycle regulation protein-3; tissue expression; cancer; | |
| KW | inflammation; phlebotomy vector; gene therapy; disease diagnosis; | |
| KW | disease monitoring; chromosomal mapping; cell proliferation; | |
| KW | cell proliferative disorder; ds. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| FT | key | Location/Qualifiers |
| FT | CDS | 140..850 |
| FT | | /*tag= a |
| FT | | /product= "Cell cycle regulation protein-3" |
| PN | MO9964593-A2. | |
| XX | | |
| PD | 16-DEC-1999. | |
| XX | | |
| PF | 08-JUN-1999; | 99WO-US12906. |
| XX | | |
| PR | 08-JUN-1999; | 98US-0088695. |
| XX | | |
| PA | (INCY-) INCYTE PHARM INC. | |

DNA representing 1
DNA encoding a p16
Human molecule for
Human CDNA 5'-end
Human CDNA clone r
Incyle clone 22384
EST clone G21. Ho
Human secreted pro
Porcine muscular s
Human polynucleoti
Incyle clone 13520
Incyle clone 30157
Novel murine polyn
Human spliced tran
Mouse spliced tran
Human NOV34a. CDNA.
Human CDNA clone (
Human CHD protein
Human haematopoiet
Human CDNA coexpre
Human polynucleoti
Human polynucleoti
Cardiac adenyllyl c
Human ovarian carc
Human ovarian carc
Ovarian carcinoma
Lung cancer associ
Human colon cancer
Cardiac adenyllyl c
GDF-1a cDNA sequen
Mouse Gli1 cell 1
Streptomycies albi
Mouse haematopoiet
Human breast cance
Streptomycies timos
S. mutrayaensis A
Streptomycies murray

| | |
|----|---|
| PI | Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR; |
| PI | Patterson C; |
| XX | WPI: 2000-105887/09. |
| DR | P-PSDB; AAY44362. |
| XX | |
| PT | Novel regulatory proteins, for diagnosis, treatment and prevention of |
| PT | cell proliferative and immune system diseases |
| XX | Claim 7; Page 75; 88bp; English. |
| XX | |
| CC | The present sequence encodes cell cycle regulation protein-3 (CECRP-3). |
| CC | Tissue expression is seen in reproductive (0.302), nervous (0.163) and |
| CC | cardiovascular (0.186) tissues. The cDNA is cloned into pBluescript |
| CC | vector. Diseases associated with tissue expression of CECRP-3 are cancer, |
| CC | inflammation and other diseases. This is used in the recombinant |
| CC | production of CECRP-3, in gene therapy (including expression of antisense |
| CC | sequences, triplex-forming molecules or ribozymes), and as source of |
| CC | probes and primers for diagnosis and monitoring of disease. It is also |
| CC | used for chromosomal mapping and isolation of related sequences. CECRPs |
| CC | are activators of cell proliferation or inhibitors of cellular processes |
| CC | that modulate proliferation. They are used to treat cell proliferative |
| CC | disorders. |
| XX | |
| SQ | Sequence 1192 BP; 237 A; 350 C; 375 G; 230 T; 0 other; |
| | |
| | Query Match 100.0%; Score 1192; DB 21; Length 1192; |
| | Best Local Similarity 100.0%; Pred. No. 1,2e-293; |
| | Matches 1192; Conservative 0; Mismatches 0; Indels 0; Gaps 0 |
| OY | 1 GTGAGCCGGAGTCAAGAACTGGCTCTCCGGACCCAGGCGCGGTTTCCGAGACAGCCAA 60 |
| DB | 1 GTGAGCCGGAGTCAAGAACTGGCTCTCCGGACCCAGGCGCGGTTTCCGAGACAGCCAA 60 |
| OY | 61 CAACGATGCTGCGCGCGCGCTTTCATTGTTGGTGGTGGTACCTCTTGCTTGA 120 |
| DB | 61 CAACGATGCTGCGCGCGCGCTTTCATTGTTGGTGGTGGTACCTCTTGCTTGA 120 |
| OY | 121 TTGGCCGCTAGTACAAAGATGCTGAGCAAGGCTTGAACCGGAAACGGAGAGAGGA 180 |
| DB | 121 TTGGCCGCTAGTACAAAGATGCTGAGCAAGGCTTGAACCGGAAACGGAGAGAGGA 180 |
| OY | 181 GGAAGAAGAACTCTGGGAGTCACTCTGGTGGCTTAATCTCGACACGAGCGGTGC 240 |
| DB | 181 GGAAGAAGAACTCTGGGAGTCACTCTGGTGGCTTAATCTCGACACGAGCGGTGC 240 |
| OY | 241 ACAGGCAACCCGCGCGCTGAGCTTACCTCTTTGACCTCTGAGTGCCTCAAGCTCA 300 |
| DB | 241 ACAGGCAACCCGCGCGCTGAGCTTACCTCTTTGACCTCTGAGTGCCTCAAGCTCA 300 |
| OY | 301 CCACAGCCTGAGAGAGTGAAGCCGGAACCTGCGGCACTGGTGGTGTGAGACACT 360 |
| DB | 301 CCACAGCCTGAGAGAGTGAAGCCGGAACCTGCGGCACTGGTGGTGTGAGACACT 360 |
| OY | 361 GCGGCGCATCCAGGCGTCCATGAGCAACCGCGGCTGCCCTGACCTGTGCTAGCCAC 420 |
| DB | 361 GCGGCGCATCCAGGCGTCCATGAGCAACCGCGGCTGCCCTGACCTGTGCTAGCCAC 420 |
| OY | 421 TGCAAGCCCCAGTGTGGCTGACAACTTACTGGAAGTCTGGAACGCTGCCCTTTCAAGCTC 480 |
| DB | 421 TGCAAGCCCCAGTGTGGCTGACAACTTACTGGAAGTCTGGAACGCTGCCCTTTCAAGCTC 480 |
| OY | 481 CATGGCCAGCCTCTGAGAGACTCAGCAATTTAGAGGCGCTGAAGTCCGCCAAC 540 |
| DB | 481 CATGGCCAGCCTCTGAGAGACTCAGCAATTTAGAGGCGCTGAAGTCCGCCAAC 540 |
| OY | 541 CTGGGACAGAGAGGGGCAACAGGCGGTAGCATGGGGAGACAGCGCCAGCTGGTGC 600 |
| DB | 541 CTGGGACAGAGAGGGGCAACAGGCGGTAGCATGGGGAGACAGCGCCAGCTGGTGC 600 |
| OY | 601 CTGGACCTCTGGGCGGCAAGCAATGGTGTCTACTGACATGGGCTTGAAGGCTGT 660 |
| DB | 601 CTGGACCTCTGGGCGGCAAGCAATGGTGTCTACTGACATGGGCTTGAAGGCTGT 660 |

| | | | | |
|-----------------------------------|--|------|--|------|
| D | b | 601 | CTTGGACCTGCTGGGCGACGCACTGGCTGCTCTACTGAGCATGGGCTTGAGGCGCTGTT | 660 |
| Q | y | 661 | TGAGGATATTGACACCCCTCTATGTTATGACAAATGAACCTTTGGGACACCACCTCTGAGGCGCT | 720 |
| D | b | 661 | TTGAGGATATTGACACCCCTCTATGTTATGACAAATGAACCTTTGGGACACCACCTCTGAGGCGCT | 720 |
| Q | y | 721 | CAAAACGAGGCCCTTAGAGATGGGCGGGCCAAAGSAGAAAGCTCCGAGCTGGACAGAGCGCGA | 780 |
| D | b | 721 | CAAAACGAGGCCCTTAGAGATGGGCGGGCCAAAGSAGAAAGCTCCGAGCTGGACAGAGCGCGA | 780 |
| Q | y | 781 | ATTGGACTACCTCATGAGATGCTGCTGGGCGACACAGGCACTGGAGCGACCGCGGGGCC | 840 |
| D | b | 781 | ATTGGACTACCTCATGAGATGCTGCTGGGCGACACAGGCACTGGAGCGACCGCGGGGGCC | 840 |
| Q | y | 841 | AGGCGGCTGAGACCCCTCGTGCCTGAGATGGTGTCTGATATGCAACTGAGCCTGTGCTG | 900 |
| D | b | 841 | AGGCGGCTGAGACCCCTCGTGCCTGAGATGGTGTCTGATATGCAACTGAGCCTGTGCTG | 900 |
| Q | y | 901 | GACCAACTGTCTTGAAAAGACACAGCTGGCTTCCCTAGTACAGAGAACAGGGCTTGGGC | 960 |
| D | b | 901 | GACCAACTGTCTTGAAAAGACACAGCTGGCTTCCCTAGTACAGAGAACAGGGCTTGGGC | 960 |
| Q | y | 961 | CACTTTGGAGAGACAGATCTAGTCTGGGCAACTTACATCCGTCCTCTGTCTCAGGG | 1020 |
| D | b | 961 | CACTTTGGAGAGACAGATCTAGTCTGGGCAACTTACATCCGTCCTCTGTCTCAGGG | 1020 |
| Q | y | 1021 | CTGGCAGGGGAGCCCTGGAATTACCCCTAGTGAATGACAGGGCTGTGGGAGCT | 1080 |
| D | b | 1021 | CTGGCAGGGGAGCCCTGGAATTACCCCTAGTGAATGACAGGGCTGTGGGAGCT | 1080 |
| Q | y | 1081 | GAATTCCTCGGCCCTGGGGGTCATAGCTTGGGCTGCTCTCTGATACGGAGAAAGAC | 1140 |
| D | b | 1081 | GAATTCCTCGGCCCTGGGGGTCATAGCTTGGGCTGCTCTCTGATACGGAGAAAGAC | 1140 |
| Q | y | 1141 | CCAATCAGATTTTTCAAATTAAGCCAGCTCTGGGAATGTCAAAAA | 1192 |
| D | b | 1141 | CCAATCAGATTTTTCAAATTAAGCCAGCTCTGGGAATGTCAAAAA | 1192 |
| RESULT 2 | | | | |
| AAH64895 standard; cDNA; 1219 BP. | | | | |
| AC | AAH64895: | | | |
| XX | | | | |
| DT | 11-SEP-2001 (first entry) | | | |
| XX | | | | |
| DE | Human secreted protein cDNA, SEQ ID NO: 171. | | | |
| XX | | | | |
| KM | Human; secreted protein; gene therapy; vaccine; treatment; diagnosis; | | | |
| XX | GENSET; ss. | | | |
| OS | Homo sapiens. | | | |
| XX | | | | |
| XX | WO200142451-A2. | | | |
| PD | 14-JUN-2001. | | | |
| XX | | | | |
| PF | 07-DEC-2000; 2000MO-IB01938. | | | |
| XX | | | | |
| PR | 08-DEC-1999; 9905-0169629. | | | |
| XX | | | | |
| PR | 06-MAR-2000; 2000US-0187470. | | | |
| XX | | | | |
| PA | (GENSET) GENSET. | | | |
| XX | | | | |
| PI | Dumas Milne Edwards J, Bougueleret L, Jobert S; | | | |
| XX | | | | |
| DR | WPI; 2001-367870/38. | | | |
| XX | | | | |
| DR | P-PSDB; AAC89292. | | | |
| PT | Full length GENSET human nucleic acids encoding potentially secreted proteins, useful in gene therapy and vaccination against a variety of diseases, and for diagnosis of those diseases - | | | |

XX Claim 7; Page 733-734; 921pp: English.
CC The invention relates to full length GENSET human nucleic acids encoding
CC potentially secreted proteins. The nucleic acids and the polypeptides
CC they encode may be used in the prevention, treatment and diagnosis of
CC diseases associated with inappropriate GENSET gene expression. For
CC example, they be used to treat disorders associated with decreased
CC GENSET gene expression by rectifying mutations or deletions in a
CC patient's genome that affect the activity of GENSET or by supplementing
CC the patient's own production of GENSET polypeptides. Conversely,
CC antisense nucleic acid molecules may be administered to down regulate
CC GENSET expression by binding with the cells' own genes and preventing
CC their expression. The sense and antisense nucleic acids may also be
CC used as DNA probes in diagnostic assays to detect and quantitate the
CC presence of similar nucleic acid sequences in samples, and hence to
CC determine which patients may be in need of restorative therapy.
CC The GENSET polypeptides may be used as antigens in the production of
CC antibodies and in assays to identify modulators (agonists and
CC antagonists) of GENSET polypeptide expression and activity. The
CC present sequence is a GENSET nucleic acid of the invention.
XX
SQ Sequence 1219 BP; 249 A; 352 C; 381 G; 237 T; 0 other;

Query Match 99.7%; Score 1188.8; DB 22; Length 1219;

Best Local Similarity 99.8%; Pred. No. 7.8e-293;

Matches 1190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GTGAGCCGAGTCAAGAACTGCTCTGCGACCCAGGCGGGGTTCCGGAGACAGCCAA 60
21 GTGAGCGGAGTCAAGAACTGCTCTGCGACCCAGGCGGGGTTCCGGAGACAGCCAA 80
61 CAAGCGAGTCTCCCGCCCGCTTCCTGATTTGGTGGTGGTCACTCTTCTGTTCTGA 120
81 CAAGCGAGTCTCCCGCCCGCTTCCTGATTTGGTGGTGGTCACTCTTCTGTTCTGA 140
121 TTGGCGCGTAGTGAAGAACTGCTGAGCAAGGCTCTGAAGCGGAAACGGAGAGAGAGA 180
141 TTGGCGCGTAGTGAAGAACTGCTGAGCAAGGCTCTGAAGCGGAAACGGAGAGAGAGA 200
181 GGAGAGAGAACTCTGGAGTGAAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 240
201 GGAGAGAGAACTCTGGAGTGAAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 260
241 ACAGCGACCCCGCGCGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 300
261 ACAGCGACCCCGCGCGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 320
301 CCAGAGCGTGAAGAGTGAAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 360
321 CCAGAGCGTGAAGAGTGAAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 380
361 GCGGCGCATTCAGGCGTCCATGCGACCGCGCTGCGTCCCTGCACTGTCCTAGCCAC 420
381 GCGGCGCATTCAGGCGTCCATGCGACCGCGCTGCGTCCCTGCACTGTCCTAGCCAC 440
421 TGCAGCCCCCACTGAGTGAAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 480
441 TGCAGCCCCCACTGAGTGAAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 500
481 CATGGCAGCTCTGAGAGACTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 540
501 CATGGCAGCTCTGAGAGACTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 560
541 CTTGGCAGAGAGAGGCGCACAGGCGGTAGCATTCGCGGGGAGCAGCGCCAGCTGGTGC 600
561 CTTGGCAGAGAGAGGCGCACAGGCGGTAGCATTCGCGGGGAGCAGCGCCAGCTGGTGC 620
601 CTTGACACTGCTGGGCGCCAGCGCTGCTCTACTGAGAGATGGGCTTGAAGGCGCTGT 660
621 CTTGACACTGCTGGGCGCCAGCGCTGCTCTACTGAGAGATGGGCTTGAAGGCGCTGT 680
661 TGAGGATATTGACACTCTATGATGACAAATGATTTGGGAGACAGCTCTGAGGCGCT 720

681 TGAGGATATTGACACTCTATGATGACAAATGATTTGGGAGACAGCTCTGAGGCGCT 740
721 CAACACGAGCCCTGAGATGAGGCGGCGAAGAGAGAACTCCGAGACTGGAGAGCGCA 780
741 CAACACGAGCCCTGAGATGAGGCGGCGAAGAGAGAACTCCGAGACTGGAGAGCGCA 800
781 ATTGACATCTCATATGATGATGCTGTGGGCAACAGGCACTGAGAGCGCCGGGCGC 840
801 ATTGACATCTCATATGATGATGCTGTGGGCAACAGGCACTGAGAGCGCCGGGCGC 860
841 AGGCGCTAGCCCTGCTGCTGATGATGTTGCTGTGATGATGATGATGATGATGATG 900
861 AGGCGCTAGCCCTGCTGCTGATGATGTTGCTGTGATGATGATGATGATGATGATG 920
901 GACCAACTCTCTCGAAAAGACAGAGCTGCTCCCTAGTACAGAGAAAGGCGTTGGGC 960
921 GACCAACTCTCTCGAAAAGACAGAGCTGCTCCCTAGTACAGAGAAAGGCGTTGGGC 980
961 CACTTTGAGAGACAGAAATCTAGTCTGTGGCAACTTCACATCCGTCCTGCTCAGGG 1040
981 CACTTTGAGAGACAGAAATCTAGTCTGTGGCAACTTCACATCCGTCCTGCTCAGGG 1060
1021 CTGGCAGGGGAGCCCTGGAATTAACCCCTAGTATGATGATGATGATGATGATGATG 1080
1041 CTGGCAGGGGAGCCCTGGAATTAACCCCTAGTATGATGATGATGATGATGATGATG 1100
1081 GAATTCCTCGGCGCTGGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1140
1101 TAAATCCCTGCGCCCTGGGCTCATAGTCTGATGATGATGATGATGATGATGATGATG 1160
1141 CCAATCAGATTTTTCAAATTAAGCCAGTCTGGGAAATCTCAAAAAAAAAA 1192
1161 CCAATCAGATTTTTCAAATTAAGCCAGTCTGGGAAATCTCAAAAAAAAAA 1212
RESULT 3
AAK94674
ID AAK94674 standard; cDNA; 2113 BP.
XX
AC AAK94674;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 3676.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99UP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-52425/58.
XX
PT P-Psdb; AAM93724.
XX
PS 830 Primers useful for synthesizing full length cDNA clones and their
PS use in genetic manipulation -
PS Claim 8; SEQ ID NO 3676; 1380pp + sequence listing; English.

CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5' and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesising the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX Sequence 2113 BP, 453 A; 561 C; 578 G; 521 T; 0 other;

Query Match 99.3%; Score 1183.4; DB 22; Length 2113;
 Best Local Similarity 99.9%; Pred. No. 2.2e-291;
 Matches 1184; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGAGCCGGAGTCAGACTGCGTCTGCGGACCCAGCGCGGGTTCCGGAGACAGCCAA 60
 Db 21 GTGAGCCGGAGTCAGACTGCGTCTGCGGACCCAGCGCGGGTTCCGGAGACAGCCAA 80
 QY 61 CAAGGATGTCGCGCGCGCGTTCCTGATGTTGGTGGTGGCTACCTCTTCCTTGA 120
 Db 81 CAAGGATGTCGCGCGCGCGTTCCTGATGTTGGTGGTGGCTACCTCTTCCTTGA 140
 QY 121 TTGGCCGCTAGTAGCAAGATGCTGAGCAAGGTTCTGAAGCGAAGAGGAGAGAGA 180
 Db 141 TTGGCCGCTAGTAGCAAGATGCTGAGCAAGGTTCTGAAGCGAAGAGGAGAGAGA 200
 QY 181 GGAGAGAGAACCTCTGCGAGTCGACTCTGTGGCTAGATCTCGGACAGCGGTTGC 240
 Db 201 GGAGAGAGAACCTCTGCGAGTCGACTCTGTGGCTAGATCTCGGACAGCGGTTGC 260
 QY 241 ACAGGACACCCCGCGCGCGTCTAGTCCCTTTTACCTCTGAGTCTCAAGCTCA 300
 Db 261 ACAGGACACCCCGCGCGCGTCTAGTCCCTTTTACCTCTGAGTCTCAAGCTCA 320
 QY 301 CCACAGCCTGAGCAGAGTAGTGACGCGGACCTGCGGACCTGCTGCTGTAACACT 360
 Db 321 CCACAGCCTGAGCAGAGTAGTGACGCGGACCTGCGGACCTGCTGCTGTAACACT 380
 QY 361 GCGGCGGCTTCAGGCGTCTCATGCGACCCGCGGCTGCTGCGACCTGCTGCTGCG 420
 Db 381 GCGGCGGCTTCAGGCGTCTCATGCGACCCGCGGCTGCTGCGACCTGCTGCTGCG 440
 QY 421 TGACGCCCCCAGTGTGGCTGACAACTTACTGGAAGCTCGAGCGTCCCTTACGCTC 480
 Db 441 TGACGCCCCCAGTGTGGCTGACAACTTACTGGAAGCTCGAGCGTCCCTTACGCTC 500
 QY 481 CATGSGCCAGCCTCTGAGAGACTCAGCCACATTGAGGCGCTGAGTCAGGCTCCCAAC 540
 Db 501 CATGSGCCAGCCTCTGAGAGACTCAGCCACATTGAGGCGCTGAGTCAGGCTCCCAAC 560
 QY 541 CTTGGCAGACAGAGGCGCCACAGGCGCTGAGCATGGGGGAGACAGCGCCAGCTGGGTG 600
 Db 561 CTTGGCAGACAGAGGCGCCACAGGCGCTGAGCATGGGGGAGACAGCGCCAGCTGGGTG 620
 QY 601 CTTGGACCTGCTGGGCGCAGCAGCTGCTGCTACTGAGACATGGGCTTGAAGGCGCTG 660
 Db 621 CTTGGACCTGCTGGGCGCAGCAGCTGCTGCTACTGAGACATGGGCTTGAAGGCGCTG 680
 QY 661 TGAGGATATTGACACTCTATGATGACAAATGAATTTGGGACACAGCTTGAAGGCT 720
 Db 681 TGAGGATATTGACACTCTATGATGACAAATGAATTTGGGACACAGCTTGAAGGCT 740
 QY 721 CAACACAGGCGCTTGAAGATGGGCGGCGAAGAGAAAGCTCCGAGCTGACGAGCGCGA 780
 Db 741 CAACACAGGCGCTTGAAGATGGGCGGCGAAGAGAAAGCTCCGAGCTGACGAGCGCGA 800
 QY 781 ATTGAGCTACCTGATGATGATGCTGCTGGGCGACACAGCACTTGAAGAGACCGCGGGGCC 840
 Db 801 ATTGAGCTACCTGATGATGCTGCTGGGCGACACAGCACTTGAAGAGACCGCGGGGCC 860

Db 801 ATTGAGCTACCTGATGATGCTGCTGGGCGACACAGCACTTGAAGAGACCGCGGGGCC 860
 QY 841 AGGCGCTGAGCCCTCGTCTGATGATGCTGATGATGATGATGATGATGATGATGATG 900
 Db 861 AGGCGCTGAGCCCTCGTCTGATGATGATGATGATGATGATGATGATGATGATGATG 920
 QY 901 GACCACTGCTCTGAGAAAGACACAGCTGCTTCCCTAGTACAGAGAGAGAGAGAGAG 960
 Db 921 GACCACTGCTCTGAGAAAGACACAGCTGCTTCCCTAGTACAGAGAGAGAGAGAGAG 980
 QY 961 CACTTTGAGAGACAGATAGTCCCTGGGAGACCTGACATCCGCTCCCTGCTCAGGG 1020
 Db 981 CACTTTGAGAGACAGATAGTCCCTGGGAGACCTGACATCCGCTCCCTGCTCAGGG 1040
 QY 1021 CTGGCAGGAGGAGCCTGGAATTACCCCTAGTATGATGATGATGATGATGATGATG 1080
 Db 1041 CTGGCAGGAGGAGCCTGGAATTACCCCTAGTATGATGATGATGATGATGATGATG 1100
 QY 1081 GAATTCCTGCGCCCTGGGCTATGACCTTGGCTGCTGCTGCTGCTGCTGCTGCTG 1140
 Db 1101 GAATTCCTGCGCCCTGGGCTATGACCTTGGCTGCTGCTGCTGCTGCTGCTGCTG 1160
 QY 1141 CCAATCAGATTTTCAATTAAGCCAGTCTGGGAAATCTCAAA 1185
 Db 1161 CCAATCAGATTTTCAATTAAGCCAGTCTGGGAAATCTCAAA 1205

RESULT 4
 AAH99491/C
 ID AAH99491 standard; cDNA: 1181 BP.
 XX
 AC AAH99491;
 XX
 DT 16-OCT-2001 (first entry)
 XX
 DE Human protein encoding cDNA sequence SEQ ID NO:326.
 XX
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimetagen; cardiovascular; antianaemic; anaemia;
 KW antiagregant; haemostatic; vulnary; antitumor; osteopathic; eczema;
 KW dermatological; antiallergic; antiautismic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anapylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder; ss.
 KW
 KW Homo sapiens.
 OS
 XX WO200153455-A2.
 PN 26-JUL-2001.
 PD
 XX 22-DEC-2000; 2000WO-US35017.
 XX
 XX 23-DEC-1999; 99US-0471275.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 XX
 PA (HSE-) HSEQ INC.
 XX
 PI Tang YT, Lin C, Drmanac RT;
 XX
 DR WPI: 2001-457603/49.
 DR P-PSDB; AAM25550.
 XX
 PT Isolated human polynucleotides encoding polypeptides, useful for the

PR 31-MAR-1999; 9905-0127607.
 PR 02-APR-1999; 9905-0127636.
 PR 05-APR-1999; 9905-0127728.
 PR 30-MAR-2000; 200005-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB42418.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 3559; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferic; antiparkinsonian; nootropic; neuroprotective;
 CC osteoprotic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertoid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX
 SQ Sequence 1222 BP; 229 A; 370 C; 350 G; 270 T; 3 other:
 Query Match 96.8%; Score 1154.4; DB 21; Length 1222;
 Best Local Similarity 99.7%; Pred. No. 4-5e-284;
 Matches 1178; Conservative 0; Mismatches 1; Indels 3; Gaps 2.
 QY 14 AGAAGTGGTCTGCGACCAAGGCGGCGGTTTCGGAGACAGCCAAAGCATGTGTC 73
 DB 1217 AGAAGTGGTCTGCGACCAAGGCGGCGGTTTCGGAGACAGCCAAAGCATGTGTC 1158
 QY 74 CGCGCGCTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 133
 DB 1157 CGCGCGCTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATG 1098
 QY 134 AGCAAGATGCTGACCAAGGCTCTGAAGCGGAACGAGGAGAGAGAGAGAGAGAACT 193
 DB 1097 AGCAAGATGCTGACCAAGGCTCTGAAGCGGAACGAGGAGAGAGAGAGAGAACT 1038
 QY 194 CTGGCAGTCACTCCTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 253
 DB 1037 CTGGCAGTCACTCCTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 978
 QY 254 GCGGCGCTTCTAGTCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 312
 DB 977 GCGGCGCTTCTAGTCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 918
 QY 313 GCAGAGTACGCGGACCTGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 372
 DB 917 GCAGAGTACGCGGACCTGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 858
 QY 373 --GGGTCATGAGACCGCGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 430
 DB 857 NNGGCTGTCATGAGACCGCGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 798

QY 431 AGTGTGCTGACAACTTACTGCAAGCTCGGACGCTGCCCTTTCAGCCTTCATGCGCAGC 490
 DB 797 AGTGTGCTGACAACTTACTGCAAGCTCGGACGCTGCCCTTTCAGCCTTCATGCGCAGC 738
 QY 491 CTCTGAGAGACCTTACGACATTTGAGGCGCTGAGTACGCTGCCCAACCTTTGGCAGAC 550
 DB 737 CTCTGAGAGACCTTACGACATTTGAGGCGCTGAGTACGCTGCCCAACCTTTGGCAGAC 678
 QY 551 GAGGAGCCACAGGCGCTGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 610
 DB 677 GAGGAGCCACAGGCGCTGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 618
 QY 611 CTGGGCGGACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 670
 DB 617 CTGGGCGGACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 558
 QY 671 GACACCTTATGATGACAAATGAACTTTGGGACACGAGCTGCTGAGGCGCTTAAACAGGC 730
 DB 557 GACACCTTATGATGACAAATGAACTTTGGGACACGAGCTGCTGAGGCGCTTAAACAGGC 498
 QY 731 CCTGAGATGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 790
 DB 497 CCTGAGATGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 438
 QY 791 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 850
 DB 437 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 378
 QY 851 GCGGCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 910
 DB 377 GCGGCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 318
 QY 911 CCTGAGAAAGACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 970
 DB 317 CCTGAGAAAGACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 258
 QY 971 AGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1030
 DB 257 AGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 198
 QY 1031 GAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1090
 DB 197 GAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 138
 QY 1091 GCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1150
 DB 137 GCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 78
 QY 1151 TTTTCAAATTAAGCCAGTCTGAGGAAATCTCAAAAAAAA 1192
 DB 77 TTTTCAAATTAAGCCAGTCTGAGGAAATCTCAAAAAAAA 36
 RESULT 6
 AAX97688
 ID AAX97688 standard; DNA; 1182 BP.
 AAX97688:
 DT 13-SEP-1999 (first entry)
 XX
 DE Extended human secreted protein coding sequence, SEQ ID NO. 253.
 XX
 KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
 KW cellular differentiation; immune system regulator; anti-inflammatory;
 KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
 KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
 KW genetic disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN W09931236-A2.

XX 24-JUN-1999.
 PD 17-DEC-1998; 98WO-1B02122.
 XX 10-AUG-1998; 98US-0096116.
 PR 17-DEC-1997; 97US-0069957.
 PR 09-FEB-1998; 98US-0074121.
 PR 13-APR-1998; 98US-0081563.
 XX (BEST) GENSEFT.
 PA Bougueleret L, Duclert A, Dumas Milne Edwards J;
 PI WPL; 1999-385906/32.
 DR P-PSDB; AAY36004.
 XX
 PT New Isolated human secreted proteins
 PS Claim 1; Page 330-331; 516pp; English.
 CC This sequence represents an extended human secreted protein coding
 CC sequence of the invention. The secreted proteins can be used in treating
 CC or controlling a variety of human conditions. The secreted proteins may
 CC act as cytokines or may affect cellular proliferation or differentiation
 CC or may act as immune system regulators, haematopoiesis regulators, tissue
 CC growth regulators, regulators of reproductive hormones or cell movement
 CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
 CC tumour inhibition activity. The DNAs can be used in forensic procedures
 CC to identify individuals or in diagnostic procedures to identify
 CC individuals having genetic diseases resulting from abnormal expression of
 CC the genes corresponding to the extended cDNAs. They are also useful for
 CC constructing a high resolution map of the human chromosomes. They can
 CC also be used for gene therapy to control or treat genetic diseases.
 CC
 XX Sequence 1182 BP; 239 A; 348 C; 357 G; 227 T; 11 other;
 SQ
 Query Match 96.6%; Score 1151.2; DB 20; Length 1182;
 Best Local Similarity 98.6%; Pred. No. 2,9e-283;
 Matches 1162; Conservative 8; Mismatches 8; Indels 1; Gaps 1;

QY 494 CTGGAGAGCTCAGCCACATTTAGAGGCGTAGTCAGGCTCCCAACCCCTTGGCAGACGAG 553
 DB 482 CTGGARAGACCTCAGCCACATTTAGAGGCGCTAGTCAGGCTCCCAACCCCTTGGCAGACGAG 541
 QY 554 GGGCCACACAGCCGCTACATTCGGGGAGCAGCCGCCAGCCTGGTGGCTTGGACCTGCTG 613
 DB 542 GGGCCACACAGCCGCTACATTCGGGGAGCAGCCGCCAGCCTGGTGGCTTGGACCTGCTG 601
 QY 614 GGGCCACACAGCCGCTACATTCGGGGAGCAGCCGCCAGCCTGGTGGCTTGGACCTGCTG 673
 DB 602 GGGCCACACAGCCGCTACATTCGGGGAGCAGCCGCCAGCCTGGTGGCTTGGACCTGCTG 661
 QY 674 ACCCTATGTATGACATTAATGAACTTTGGGACACAGCCCTGTAGGCGCTCAACACAGGCGCT 733
 DB 662 ACCCTATGTATGACATTAATGAACTTTGGGACACAGCCCTGTAGGCGCTCAACACAGGCGCT 721
 QY 734 GAGGATGGGCGGGCAAGAGAAAGCTCCGAGCTGAGACAGAGCCGAATTTGACTACTTC 793
 DB 722 GAGGATGGGCGGGCAAGAGAAAGCTCCGAGCTGAGACAGAGCCGAATTTGACTACTTC 781
 QY 794 ATGGATGTGCTGGTGGGCAACAGAGCACTGGAGCAGCCGCGGCGCCAGAGCGCTGAGCC 853
 DB 782 ATGGATGTGCTGGTGGGCAACAGAGCACTGGAGCAGCCGCGGCGCCAGAGCGCTGAGCC 841
 QY 854 CTGCTGTGGAAATGTTGTGTGTATGTAAGTGAAGCTGCTGCTGGACCAACTGCTCT 913
 DB 842 CTGCTGTGGAAATGTTGTGTGTATGTAAGTGAAGCTGCTGCTGGACCAACTGCTCT 901
 QY 914 CGAAAGACACAGCTGGCTTCCCTAGTACAGAGAACAGGCGCTTGGCGCACTTTGAGAGAG 973
 DB 902 CGAAAGACACAGCTGGCTTCCCTAGTACAGAGAACAGGCGCTTGGCGCACTTTGAGAGAG 961
 QY 974 CAGATGTACTCCGGGCACTTCACTACCTGCTCTCTGCTCAAGGCTGGCGAGGGGAG 1033
 DB 962 CAGATGTACTCCGGGCACTTCACTACCTGCTCTCTGCTCAAGGCTGGCGAGGGGAG 1021
 QY 1034 CCTGGAATTAACCCCTAGTATGATGAGAGGAGTGTGGGAGTGAATTCCTGGCC 1093
 DB 1022 CCTGGAATTAACCCCTAGTATGATGAGAGGAGTGTGGGAGTGAATTCCTGGCC 1081
 QY 1094 CTGGGCTCATAGCTTGGGCTGTTCTTCTGTGATGCGGGAAGACCCCAATCAGATTTT 1153
 DB 1082 CTGGGCTCATAGCTTGGGCTGTTCTTCTGTGATGCGGGAAGAG-CCCAATCAGATTTT 1140
 QY 1154 TCAATTTAAAGCCAGTCTCGGAAATCTCAAAAAAAA 1192
 DB 1141 TCAATTTAAAGCCAGTCTCGGAAATCTCAAAAAAAA 1179

RESULT 7
 AAZ19894
 ID AAZ19894 standard; cDNA; 1078 BP.
 AC AAZ19894;
 DT 06-DEC-1999 (first entry)
 XX
 XX Human adult blood secreted protein g2L_1 cDNA.
 DE
 XX Secreted protein; g2L_1; human; therapy; diagnosis; vaccine;
 KW blood; ss.
 KM
 XX Homo sapiens.
 OS
 XX
 XX
 FH Key location/Qualifiers
 FT CDS 10..720
 FT /tag= a
 FT /note= "this region is specifically claimed in
 FT Claim 12(b)"
 FT sig_peptide 10..246
 FT mat_peptide 247..717
 FT /tag= b
 FT /tag= c

```

FT      /note= "this region (plus the stop codon) is
FT      sig_peptide 10..102 specifically claimed in Claim 12(c) "
FT      /*tag= d
FT      /note= "alternative signal sequence"
FT      mat_peptide 103..717
FT      /*tag= e
FT      /note= "alternative mature protein"
XX      WO947555-A1.
XX      23-SEP-1999.
XX      18-MAR-1999; 99WO-US05939.
XX      20-MAR-1998; 98US-0078803.
XX      17-MAR-1998; 98US-0078803.
XX      (GENY ) GENETICS INST INC.
XX      Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
XX      Merberg D, Treacy M, Agostino MJ, Steinhinger RJ;
XX      P-PSDB; AAY31829.
XX      WPI: 1999-562059/47.
XX      P-PSDB; AAY31829.
XX      New polynucleotides derived from murine foetal cell cDNA libraries,
XX      potentially used as, e.g. vaccines -
XX      Claim 12(a); Page 93; 107pp; English.
XX
XX      This is the nucleotide sequence of an isolated cDNA clone that
XX      codes for a novel human secreted protein, g21_1 (see AAY31829).
XX      The clone was isolated from a human adult blood (peripheral blood
XX      mononuclear cells treated with phytohemagglutinin, phorbol
XX      myristate acetate and mixed lymphocyte reaction) cDNA library
XX      using human cDNAs that had themselves been identified in a human
XX      adult blood cDNA library using methods selective for cDNAs encoding
XX      secreted proteins. The clone is deposited in composite deposit
XX      ATCC 99700. The invention provides new human secreted proteins
XX      (see AAY31828-38) and polynucleotides (see AAY19893-901) isolated from
XX      foetal cell, adult blood, adult brain and foetal kidney cDNA
XX      libraries. They are predicted to have biological activities which
XX      would make them suitable for treating, preventing or ameliorating
XX      medical conditions in humans and animals, although no supporting
XX      data are given. Suggested activities include nutritional,
XX      cytokine, tissue growth, cell proliferation and differentiation,
XX      immunostimulant (e.g. as vaccine), immunosuppressive,
XX      haematopoiesis regulating, activin or inhibin, chemotactic or
XX      chemokinetic, haemostatic or thrombolytic, receptor/ligand activity,
XX      antiinflammatory, cadherin or tumour invasion suppressor, and
XX      tumour inhibitory activities. The polynucleotides are also stated
XX      to be useful for gene therapy.
XX
XX      Sequence 1078 BP: 234 A; 311 C; 332 G; 199 T; 2 other:
SQ
Query Match 89.0%; Score 1061.2; DB 20; Length 1078;
Best Local Similarity 99.8%; Pred. No. 2.4e-260;
Matches 1060; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      |||||
181 CAGCAGATGAGCGGACCTCGGACCTGGTGTGGTGGTGAACACTCTCGGGGCATC 240
QY      CAGGCTCCATAGGCACCCGCGGCTGCTCGACACCTGGCCATGAGCCACATCGAGCCCC 430
Db      |||||
241 CAGGCTCCATAGGCACCCGCGGCTGCTCGACACCTGGCCATGAGCCACATCGAGCCCC 300
QY      AGTGTGCTGACAACTACTGCGCAAGCTCGGACGCTGCCCTTTCAGCTTCATGGCCAGC 490
Db      |||||
301 AGTGTGCTGACAACTACTGCGCAAGCTCGGACGCTGCCCTTTCAGCTTCATGGCCAGC 360
QY      CTCTGAGAGACCTCAGCCACATTGAGGCGCTGAGTCAAGGCTCCCAACCTTGGCAGAC 550
Db      |||||
361 CTCTGAGAGACCTCAGCCACATTGAGGCGCTGAGTCAAGGCTCCCAACCTTGGCAGAC 420
QY      GAGGGGCCACACAGGCGCTGAGTCAAGGCGCGGAGAGAGCCAGCCAGCTGGGCTGGAGCTG 610
Db      |||||
421 GAGGGGCCACACAGGCGCTGAGTCAAGGCGCGGAGAGAGCCAGCCAGCTGGGCTGGAGCTG 480
QY      CTGGGCCACGCACTGCTGTCTACTGAGCAGATGGGCTTGAAGGCTTTGAGGATATT 670
Db      |||||
481 CTGGGCCACGCACTGCTGTCTACTGAGCAGATGGGCTTGAAGGCTTTGAGGATATT 540
QY      GACACCTCTATGTATGACAAATGAACTTGGGACACAGCTCTGAGGGGCTCAAAACAGGC 600
Db      |||||
601 GACACCTCTATGTATGACAAATGAACTTGGGACACAGCTCTGAGGGGCTCAAAACAGGC 730
QY      CCTGAGATGGGCGCGGACAGAGAGAGCTCGGAGCTGGAGAGCGGAGGAGGAGGAGGAGGAG 790
Db      |||||
791 CCTGAGATGGGCGCGGACAGAGAGAGCTCGGAGCTGGAGAGCGGAGGAGGAGGAGGAGGAG 660
QY      CTGATGATGTGCTGTGGTGGGACACAGGCACTGAGCAGCAGCCGCGGCGCAGGCGCTGA 850
Db      |||||
851 CTGATGATGTGCTGTGGTGGGACACAGGCACTGAGCAGCAGCCGCGGCGCAGGCGCTGA 720
QY      GCGCTGCTGTGGAATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 910
Db      |||||
911 GCGCTGCTGTGGAATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
QY      CCTGGAAGACACAGCTGCGCTTCCCTAGTACAGAGAGAGGAGGAGGAGGAGGAGGAGGAG 970
Db      |||||
781 CCTGGAAGACACAGCTGCGCTTCCCTAGTACAGAGAGAGGAGGAGGAGGAGGAGGAGGAG 840
QY      AGACGAAATCTAGTCTGTGGGCAACTTCACATCCGCTCTGTCTGACAGGCTGGGAGGCG 900
Db      |||||
901 GAGCTGGAATTCACCCCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1090
QY      GCGCTGGAATTCACCCCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db      |||||
961 GCGCTGGAATTCACCCCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY      TTTTCAATTTAAGCAGTCTCGGGAATTCACCAAAAAA 1192
Db      |||||
1021 TTTTCAATTTAAGCAGTCTCGGGAATTCACCAAAAAA 1062

```

RESULT 8
AAS78542/c
ID AAS78542 standard; cDNA, 1216 BP.
XX AAS78542;
AC AAS78542;
XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #14346.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX

Query Match 59.4%; Score 707.8; DB 20; Length 711;
 Best Local Similarity 99.7%; Pred. No. 2.2e-170;
 Matches 709; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

140 ATGCTAGCAAGAGGTGTGAAGCGGAACGGAGAGAGAGAGAGAGAGAACTCTGGCA 199
 |||||||
 1 ATCTTAGCAAGAGGTGTGAAGCGGAACGGAGAGAGAGAGAGAGAGAACTCTGGCA 60
 |||||||
 200 GTGCACTCTGTGTGTAGATCTGTGGCAAGCAGCGGTGGACACAGCCCGGGCGCTG 259
 |||||||
 61 GTGCACTCTGTGTGTAGATCTGTGGCAAGCAGCGGTGGACACAGCCCGGGCGCTG 120
 |||||||
 260 GCTCTAGCTCTCTCTTTAGCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 319
 |||||||
 121 GCTCTAGCTCTCTCTTTAGCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 180
 |||||||
 320 GAGCCGAGCTGTGGGAGCTGTGGTGTGGTGTGAACACTGTGGGGGCAATCCAGGGGTCC 379
 |||||||
 181 GAGCCGAGCTGTGGGAGCTGTGGTGTGGTGTGAACACTGTGGGGGCAATCCAGGGGTCC 240
 |||||||
 380 ATGGCAACCGCGGCTGCTGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 439
 |||||||
 241 ATGGCAACCGCGGCTGCTGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 300
 |||||||
 440 GACAACTTACTGGCAAGCTGTGAGCTGTGGCTTTGAGTCTGAGTCTGAGTCTGAGTCTG 499
 |||||||
 301 GACAACTTACTGGCAAGCTGTGAGCTGTGGCTTTGAGTCTGAGTCTGAGTCTGAGTCTG 360
 |||||||
 500 GACCTAGCCACATTTAGGAGCTGTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 559
 |||||||
 361 GACCTAGCCACATTTAGGAGCTGTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 420
 |||||||
 560 CCAGGCGGTAGCATCGGGGAGAGAGCGCCAGCCTGTGGTGTGGTGTGGTGTGGTGTGGT 619
 |||||||
 421 CCAGGCGGTAGCATCGGGGAGAGAGCGCCAGCCTGTGGTGTGGTGTGGTGTGGTGTGGT 480
 |||||||
 620 GCCACGT 679
 |||||||
 481 GCCACGT 540
 |||||||
 680 ATGTATGACATTAATTTGGGACACAGCTGTGAGGCTGCTGAAACAGAGCCCTGAGAT 739
 |||||||
 541 ATGTATGACATTAATTTGGGACACAGCTGTGAGGCTGCTGAAACAGAGCCCTGAGAT 600
 |||||||
 740 GGGCGGGAAGAGAGAGCTCGGAGCTGTGAGAGAGCGCAATTTGACTTACCTCATGAT 799
 |||||||
 601 GGGCGGGAAGAGAGAGCTCGGAGCTGTGAGAGAGCGCAATTTGACTTACCTCATGAT 660
 |||||||
 800 GT 850
 |||||||
 661 GT 711
 |||||||

RESULT 11
 AAL49936 standard; cDNA; 5869 BP.
 AAL49936;
 10-DEC-2002 (first entry)

Human molecule for disease detection and treatment coding sequence #10.
 Human; molecule for disease detection and treatment; MDR; gene therapy;
 cytostatic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;
 antiinflammatory; antiasomatic; cerebroprotective; nootropic;
 neuroprotective; antiparkinsonian; cardiac; antilanginal; gene; ss.
 Homo sapiens.
 OS
 XX
 PN WO200270709-A2.
 XX
 12-SEP-2002.

XX
 PF 08-FEB-2002; 2002MO-US03709.
 XX
 PR 09-FEB-2001; 2001US-268117P.
 PR 15-FEB-2001; 2001US-269618P.
 PR 23-FEB-2001; 2001US-271118P.
 PR 07-MAR-2001; 2001US-274486P.
 PR 09-MAR-2001; 2001US-274436P.
 PR 28-NOV-2001; 2001US-334229P.
 PR 01-FEB-2002; 2002US-353284P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lal PG, Baughn MR, Yao MG, Walla NK, Elliot VS, Xu Y;
 PI Honcheil CD, Yue H, Ding L, Gietzen KU, Ison CH, Lu DAM;
 PI Hafalia AJA, Ghadi AR, Thangavelu K, Sanjanna MM, Tang YT;
 PI Ramkumar J, Griffin JA, Swannaker A, Azimati Y, Sapperstein SK;
 PI Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;
 XX
 DR MPI; 2002-713453/77.
 DR P-PSDB; AAO19407.
 XX
 PT New human molecules for disease detection and treatment (MDR), useful
 PT for diagnosing, treating and preventing diseases or conditions
 PT associated with the aberrant MDR expression, e.g. cancer, AIDS,
 PT asthma, diabetes, hepatitis -
 XX
 PS Claim 11; Page 166-167; 177pp; English.
 XX
 CC The present invention relates to human proteins and coding sequences of
 CC molecules for disease detection and treatment MDR. The sequences can be
 CC used in the treatment of diseases associated with the decreased
 CC expression or overexpression of MDR, such as cell proliferative (cancer,
 CC atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,
 CC allergies, Addison's disease, asthma), developmental (dwarfism, renal
 CC tubular acidosis), neurological (e.g. stroke, Parkinson's disease,
 CC epilepsy) and cardiovascular (congestive heart failure, myocardial
 CC infarction, angina pectoris) disorders. The present sequence is a
 CC coding sequence of the invention.
 XX
 SQ Sequence 5869 BP; 1191 A; 1682 C; 1995 G; 1001 T; 0 other;

Query Match 56.0%; Score 667.4; DB 24; Length 5869;
 Best Local Similarity 99.9%; Pred. No. 8.5e-160;
 Matches 668; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

138 AGATGTGACCAAGGTCTGAAGCGGAACGGAGAGAGAGAGAGAGAACTCTG 197
 |||||||
 237 AGATGTGACCAAGGTCTGAAGCGGAACGGAGAGAGAGAGAGAACTCTG 296
 |||||||
 198 CAGTCGACCTCTGT 257
 |||||||
 297 CAGTCGACCTCTGT 356
 |||||||
 258 TGGCTCTAGCTCTCTTTGACCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 317
 |||||||
 357 TGGCTCTAGCTCTCTTTGACCTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 416
 |||||||
 318 GTGAGCGGAGACTCGCGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 377
 |||||||
 417 GTGAGCGGAGACTCGCGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 476
 |||||||
 378 CCATGACACCGGCGGTGCTGTGACCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 437
 |||||||
 477 CCATGACACCGGCGGTGCTGTGACCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 536
 |||||||
 438 CTGACAACTTACTGTGCAAGCTGTGAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 497
 |||||||
 537 CTGACAACTTACTGTGCAAGCTGTGAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 596
 |||||||
 498 AGGACTTGAGCAATTTGAGGCGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGT 557
 |||||||
 597 AGGACTTGAGCAATTTGAGGCGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGT 656
 |||||||

```

QY 558 CACGAGCCGTCAGCATCGGGGAGCAGCCAGCCGCTGCTGCTTGGACCTGCTGGCC 617
DB 657 CACGAGCCGTCAGCATCGGGGAGCAGCCAGCCGCTGCTGCTTGGACCTGCTGGCC 716
QY 618 CAGCAGCTGCTGCTTACTGACATGAGGCTTGGAGGCTTGTGAGATATGACACCT 677
DB 717 CAGCAGCTGCTGCTTACTGACATGAGGCTTGGAGGCTTGTGAGATATGACACCT 776
QY 678 CTATGATGACATGACATGCTTGGAGCAGCCGCTGAGGCTTCAACAGGCTTGAAG 737
DB 777 CTATGATGACATGACATGCTTGGAGCAGCCGCTGAGGCTTCAACAGGCTTGAAG 836
QY 738 ATGAGCCGCGGCAAGAGAGAGCTCCGAGCTGAGAGAGCGCAATTGACATGATG 797
DB 837 ATGAGCCGCGGCAAGAGAGAGCTCCGAGCTGAGAGAGCGCAATTGACATGATG 896
QY 798 ATGCTGCG 806
DB 897 ATGCTGCG 905

```

RESULT 12

AAK91707 standard; cDNA; 696 BP.

AAK91707;

06-NOV-2001 (first entry)

Human cDNA 5'-end sequence, SEQ ID NO: 167.

Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

EPI130094-A2.

05-SEP-2001.

07-JUL-2000; 2000EP-0114089.

08-JUL-1999; 99JP-0194486.

11-JAN-2000; 2000JP-0118774.

02-MAY-2000; 2000JP-0185765.

(HELI-) HELIX RES INST.

Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,

Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

WPI; 2001-524255/58.

Claim 2; SEQ ID NO 167; 1380bp + sequence listing; English.

830 Primers useful for synthesizing full length cDNA clones and their

use in genetic manipulation -

The invention relates to primers for synthesizing full length cDNA

clones. 830 cDNA molecules encoding a human protein have been

isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

molecules have been determined. Primers for synthesizing the full length

cDNA are useful for clarifying the function of the protein encoded by

the cDNA. The full length clones were obtained by construction of full

length enriched cDNA libraries that were synthesised by the oligo-capping

method. The primers enable the production of the full length cDNA easily

without any special methods. The present sequence is the nucleotide

sequence of the 5'-end of a cDNA provided in the invention.

Note: The sequence data for this patent did not form part of the printed

specification, but was obtained in CD-ROM format directly from EPO.

Sequence 696 BP; 120 A; 221 C; 220 G; 132 T; 3 other;

Query Match 53.5%; Score 637.2; DB 22; Length 696;
 Best Local Similarity 98.1%; Pred. No. 2.1e-152;
 Matches 664; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

```

QY 1 GTGAGCCGAGTCAGAACTGCGTCTCCGAGCCAGCCGCGGTTTCCGAGAGACCA 60
DB 21 GTGAGCCGAGTCAGAACTGCGTCTCCGAGCCAGCCGCGGTTTCCGAGAGACCA 80
QY 61 CAAGGATGCTGCGCCGCGCTTTCATTTGTTGGTGGTGGTACCTCTTCTTGA 120
DB 81 CAAGGATGCTGCGCCGCGCTTTCATTTGTTGGTGGTGGTACCTCTTCTTGA 140
QY 121 TTGCGCTAGTGAAGCAAGATCTGAGCAAGGCTGGAAGCGGAACGAGAGAGA 180
DB 141 TTGCGCTAGTGAAGCAAGATCTGAGCAAGGCTGGAAGCGGAACGAGAGAGA 200
QY 181 GGAGAGAGAACCTCGGAGTGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
DB 201 GGAGAGAGAACCTCGGAGTGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 260
QY 241 ACAGGACCCCGCGCGCTGAGCTGAGCTCTTGTGACCTCTCAAGTCAAGCTCA 300
DB 261 ACAGGACCCCGCGCGCTGAGCTGAGCTCTTGTGACCTCTCAAGTCAAGCTCA 320
QY 301 CCACAGCTCTGAGAGAGAGTGAAGCGGAGCTGAGCGACCTGCTGCTGCTGCT 360
DB 321 CCACAGCTCTGAGAGAGAGTGAAGCGGAGCTGAGCGACCTGCTGCTGCTGCT 380
QY 361 GCGGCGCATCCAGGCGTCCATGAGCAACCGCGGCTGCTGCTGCTGCTGCTGCT 420
DB 381 GCGGCGCATCCAGGCGTCCATGAGCAACCGCGGCTGCTGCTGCTGCTGCTGCT 440
QY 421 TGCAAGCCCGCAGTGTGCTGACAACTTACTGAGCTGAGCTGAGCTGAGCTG 480
DB 441 TGCAAGCCCGCAGTGTGCTGACAACTTACTGAGCTGAGCTGAGCTGAGCTG 500
QY 481 CATGAGCCAGCTCTCGAGAGACTGAGCAATGAGAGGCTGAGTCAAGCTCCCA 540
DB 501 CATGAGCCAGCTCTCGAGAGACTGAGCAATGAGAGGCTGAGTCAAGCTCCCA 560
QY 541 CTTGAGAGAGAGAGGCGCACAGGCGTGAATGAGAGGAGAGAGAGAGAGAGAG 600
DB 561 CTTGAGAGAGAGAGGCGCACAGGCGTGAATGAGAGGAGAGAGAGAGAGAGAG 620
QY 601 CTTGAGAGAGAGAGGCGCACAGGCGTGAATGAGAGGAGAGAGAGAGAGAG 660
DB 621 CTTGAGAGAGAGAGGCGCACAGGCGTGAATGAGAGGAGAGAGAGAGAGAG 678
QY 661 TGAGGATTTGACACCT 677
DB 679 TGAGGATTTGACACCT 695

```

RESULT 13

AAK93731 standard; cDNA; 696 BP.

AAK93731;

06-NOV-2001 (first entry)

Human cDNA clone representative sequence, SEQ ID NO: 2191.

Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

EPI130094-A2.

05-SEP-2001.

07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999: 99JP-0194486.
PR 11-JAN-2000: 2000JP-0118774.
PR 02-MAY-2000: 2000JP-0183765.
XX
XX (HELI-) HELIX RES INSR.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI: 2001-524255/58.
DR
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
XX Example 11: SEQ ID NO 2191; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 696 BP; 120 A; 221 C; 220 G; 132 T; 3 other;

Query Match 53.5%; Score 637.2; DB 22; Length 696;
Best Local Similarity 98.1%; Pred. No. 2, 1e-152;
Matches 664; Conservative 0; Mismatches 11; Indels 2; Gaps 2;
XX
XX 1 GTGAGCCGGAGTGCAGAACTGCTCTGCGACCCAGCGCGGTTTCGGAGAGACACCAA 60
DB 21 GTGAGCCGGAGTGCAGAACTGCTCTGCGACCCAGCGCGGTTTCGGAGAGACACCAA 80
XX
XX 61 CAAGCGATGCTGCGCGCGCGGTTTCTGATGTTGGTGGTGGTGGTGGTGGTGGTGGT 120
DB 81 CAAGCGATGCTGCGCGCGCGGTTTCTGATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 140
XX
XX 121 TTGGCCGCTAGTGCAGCAAGATGCTGCAAGAGGCTGCAAGCGGAAACGGAGAGAGAGA 180
DB 141 TTGGCCGCTAGTGCAGCAAGATGCTGCAAGAGGCTGCAAGCGGAAACGGAGAGAGAGA 200
XX
XX 181 GGAGAAAGAACTCTGGCAGTGCATCTCTGGTGGTGAATCTTGGCCAGCAGCGGTGGC 240
DB 201 GGAGAAAGAACTCTGGCAGTGCATCTCTGGTGGTGAATCTTGGCCAGCAGCGGTGGC 260
XX
XX 241 ACAGGACACCCCGCGCGCGGCTGAGCCCTTGGACCTCTGACCTGACCTGACCTGAC 300
DB 261 ACAGGACACCCCGCGCGCGGCTGAGCCCTTGGACCTCTGACCTGACCTGACCTGAC 320
XX
XX 301 CCACAGCCTGCAGCAGAGTGAAGCCGAGCCTGCGGCACTGTCCTGTCCTGTCCTGTC 360
DB 321 CCACAGCCTGCAGCAGAGTGAAGCCGAGCCTGCGGCACTGTCCTGTCCTGTCCTGTC 380
XX
XX 361 GGGGGCCTATCCAGGCGCTCATGGACCCGCGGCTGCTGCACTGTCCTGTCCTGTCCT 420
DB 381 GGGGGCCTATCCAGGCGCTCATGGACCCGCGGCTGCTGCACTGTCCTGTCCTGTCCT 440
XX
XX 421 TGGAGCCCGCAGTGGTGGTGCAGCACTTACCTGGCAAGCTGGAGCGCGCTTTCACCGTC 480
DB 441 TGGAGCCCGCAGTGGTGGTGCAGCACTTACCTGGCAAGCTGGAGCGCGCTTTCACCGTC 500
XX
XX 481 CATGGCCAGCCTCTGGAGAGCCTCAGCCACATTTGAGGCGCTGAGTCAAGGCTCCCAAC 540
DB 501 CATGGCCAGCCTCTGGAGAGCCTCAGCCACATTTGAGGCGCTGAGTCAAGGCTCCCAAC 560
XX
XX 541 CTTGGCAGACGAGGGGCCACACGCGCTGATCGGGGAGACAGCGCCACGCTGGGTGC 600

DB 561 CTTGGCAGACGAGGGGCCACACGCGCTGATCGGGGAGACAGCCCAAGCCTGGGTGC 620
XX
XX 601 CTTGGACCTGCTGGGGCCAGCCAGCCTGCTACTGAGAGATGGGCTTGAGGGCGCTGT 660
DB 621 CTTGGACCTGCTGGGGCCAGCCAGCCTGCTACTGAGAGATGGGCTTGAGGGCGCTGT 678
XX
XX 661 TGAGGATATTGACACCT 677
DB 679 TGAGGATATTGACACCT 695

RESULT 14
AAZ29502
ID AAZ29502 standard; DNA; 447 BP.
XX
XX AAZ29502;
AC
XX 14-MAR-2000 (first entry)
DT
XX
XX Incyte clone 2238411f6 related to cell cycle regulation proteins.
DE
XX
XX Incyte clone 2238411f6; CECRP; cell cycle regulation protein;
KW cell proliferation; cell proliferative disorder; Immune disorder; ds.
XX
XX Homo sapiens.
OS
XX
XX MO9964593-A2.
PN
XX
XX 16-DEC-1999.
PD
XX
XX 08-JUN-1999; 99WO-US12906.
PF
XX
XX 08-JUN-1998; 98US-0088695.
PR
XX
XX (INCY-) INCYTE PHARM INC.
PA
XX
XX Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR;
PI Patternson C;
PI
XX
XX WPI: 2000-105887/09.
DR
XX
XX Novel regulatory proteins, for diagnosis, treatment and prevention of
PT cell proliferative and immune system diseases -
PT
XX
XX disclosure; Page 85; 88pp; English.
PS
XX
XX The present sequence Incyte clone 2238411f6 related to cell cycle
CC regulation proteins. CECRPs are activators of cell proliferation or
CC inhibitors of cellular processes that modulate proliferation. They are
CC used to treat cell proliferative and immune disorders.
CC Note: There is no specific information regarding this sequence in the
CC specification.
XX
XX
XX Sequence 447 BP; 89 A; 120 C; 149 G; 89 T; 0 other;

Query Match 37.5%; Score 447; DB 21; Length 447;
Best Local Similarity 100.0%; Pred. No. 5e-104;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
XX 696 TTTGGGACACGAGCCTGTGAGGGCTCAAAACAGGCGCTGAGATGGCGCGGCAAGAG 755
DB 1 TTTGGGACACGAGCCTGTGAGGGCTCAAAACAGGCGCTGAGATGGCGCGGCAAGAG 60
XX
XX 756 AAGCTCCGAGAGCTGAGACGAGCGCGAATTGACTACTGATGATGCTGCTGGGCAAC 815
DB 61 AAGCTCCGAGAGCTGAGACGAGCGCGAATTGACTACTGATGATGCTGCTGGGCAAC 120
XX
XX 816 AAGCACTGAGAGCAGCGCGGCGGCGGCGCTGAGCCCTGCTGCTGGAATGTTGTCG 875
DB 121 AAGCACTGAGAGCAGCGCGGCGGCGGCGCTGAGCCCTGCTGCTGGAATGTTGTCG 180
XX
XX 876 GTATCTGAACGTAGCCTGCTGCTGAGCACTGTCCTGAAAAGACACAGCTGGCTTCC 935


```

Db      |||||
181 GATCTGACTGAGCTCTGCTGAGCACTGTCTCGAAGAAACACAGCTGCTTC 240
OY      936 CTAGTACAGAGAAAGGGCTTGGCCACTTTGGAGAGACAGATCTAGCTGGCACT 995
Db      241 CTAGTACAGAGAAAGGGCTTGGCCACTTTGGAGAGACAGATCTAGCTGGCACT 300
OY      996 TCACATCCGCTCCTCTGTCTCAGAGGCTGACAGGGGAGCCTGGAATACCCCTAGTAT 1055
Db      301 TCACATCCGCTCCTCTGTCTCAGAGGCTGACAGGGGAGCCTGGAATACCCCTAGTAT 360
OY      1056 GGAATGACAGAGGCTCTGTGTGGGAGTGAATTCCTGCGCTGGGCTCATAGCTGGGCTGT 1115
Db      361 GGAATGACAGAGGCTCTGTGTGGGAGTGAATTCCTGCGCTGGGCTCATAGCTGGGCTGT 420
OY      1116 TCCTTCTCTGATACGGGAGAGACCCC 1142
Db      421 TCCTTCTCTGATACGGGAGAGACCCC 447

RESULT 15
AAV86113/C
ID      AAV86113 standard; cDNA; 450 BP.
XX
AC      AAV86113;
XX
DT      27-APR-1999 (first entry)
XX
DE      EST clone G21.
XX
KW      Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW      tissue growth; actinin; inhibin; tumour invasion suppressor; EST; human;
KW      chemokinesis; haemostasis; gene therapy; thrombolysis;
KW      receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS      Homo sapiens.
XX
PN      WO9845435-A2.
XX
PD      15-OCT-1998.
XX
PF      10-APR-1998; 98WO-US06954.
XX
PR      10-APR-1997; 97US-0835913.
XX
PA      (GENEY ) GENETICS INST INC.
XX
PI      Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI      Racle LA, Spaulding V, Treacy M;
XX
DR      WPI, 1999-070076/06.
XX
PT      New polynucleotides encoding human secreted proteins - derived from
PT      e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX
XX      ovary, pituitary, retina and colon cDNA libraries
XX
BS      Claim 1; Page 125; 633pp; English.
XX
CC      This sequence represents an expressed sequence tag (EST), and is a
CC      polynucleotide of the invention. The polynucleotides of the invention are
CC      all secreted EST sequences isolated from a variety of human tissue
CC      sources. The EST sequences and proteins encoded by them are predicted to
CC      have useful biological activities which would make them suitable for
CC      treating, preventing or ameliorating medical conditions in humans and
CC      animals, although no supporting data is given. Suggested activities
CC      include nutritional activity, immune stimulating or suppressing activity,
CC      haematopoiesis regulating activity, tissue growth activity,
CC      actinin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC      and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC      activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC      activity. The EST sequences are also stated to be useful for gene
CC      therapy.
XX
XX

```

```

SQ      Sequence 450 BP; 86 A; 152 C; 120 G; 92 T; 0 other;
Query Match      36.4%; Score 433.8; DB 20; Length 450;
Best Local Similarity 98.4%; Pred. No. 1,2e-100;
Matches 438; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
OY      550 CGAGGGGCCACCGACCCCTAGCATCGGGGGAGACGCCGACGCTGGGTGCTTGACCT 609
Db      445 CGAGGGGCCAAGAGGCGCTAGCATCGGGGGAGACGCCGACGCTGGGTGCTTGACCT 386
OY      610 GCTGGGCCACCGACCCCTAGCATCGGGGGAGACGCCGACGCTGGGTGCTTGACCT 669
Db      385 GCTGGGCCACCGACCCCTAGCATCGGGGGAGACGCCGACGCTGGGTGCTTGACCT 326
OY      670 TGACACCTCTATGTATGACATGAATCTTGGGCACGACCTGTAGGGGCTCAACACAG 729
Db      325 TGACACCTCTATGTATGACATGAATCTTGGGCACGACCTGTAGGGGCTCAACACAG 266
OY      730 CCTGAGAGTGGCCGCGCAAGAGAAAGCTCCGGAGCTGGACGAGGCCGAATTGACTA 789
Db      265 CCTGAGAGTGGCCGCGCGCAAGAGAAAGCTCCGGAGCTGGACGAGGCCGAATTGACTA 206
OY      790 CCTATGATGTGCTGTGGGACACAGCACTGGAGCGACGCCGGGGCCAGGGCGCTG 849
Db      205 CCTATGATGTGCTGTGGGACACAGCACTGGAGCGACGCCGGGGCCAGGGCGCTG 146
OY      850 AGCCCTGCTGCTGGAATGTTGTCTGTATGTGAATGAGCTGCTGCTGGACCACTG 909
Db      145 AGCCCTGCTGCTGGAATGTTGTCTGTATGTGAATGAGCTGCTGCTGGACCACTG 86
OY      910 TCCTGGAAGAACACAGCTGGCTCCCTAGTACAGAAACAGAGGCTTGGGCACTTGA 969
Db      85 TCCTGGAAGAACACAGCTGGCTCCCTAGTACAGAAACAGAGGCTTGGGCACTTGA 26
OY      970 GAGACGAATCTAGTCTGCGCAAC 994
Db      25 GAGACGAATCTAGTCTGCTGCAC 1

```

Search completed: August 25, 2003, 13:51:51
Job time : 374 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 12:19:01 ; Search time 4428 Seconds

(without alignments)
11012.709 Million cell updates/sec

Title: US-09-701-675A-8

Perfect score: 1192
Sequence: 1 gtagccgagtcagaactg.....gggaatctcaaaaaaaaaa 1192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database : GenBank1:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pin:*
35: em_hlg_rod:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|-----------|--------------------|
| 1 | 1183.4 | 99.3 | 2113 | 6 | BD127703 | BD127703 Primer fo |
| 2 | 1183.4 | 99.3 | 2113 | 6 | AK074652 | AK074652 Homo sapi |
| 3 | 1159.9 | 97.3 | 1179 | 9 | BC002670 | BC002670 Homo sapi |
| 4 | 1159.9 | 97.2 | 1218 | 9 | AF117959 | AF117959 Homo sapi |
| 5 | 1151.2 | 96.6 | 1182 | 6 | BD139394 | BD139394 Extended |
| 6 | 1130.4 | 94.8 | 1178 | 9 | AF366402 | AF366402 Homo sapi |
| 7 | 1048 | 87.9 | 5917 | 9 | AF130860 | AF130860 Homo sapi |
| 8 | 1046.4 | 87.8 | 160643 | 9 | AC010271 | AC010271 Homo sapi |
| 9 | 777.8 | 65.3 | 5746 | 6 | AX472956 | AX472956 Sequence |
| 10 | 773.8 | 61.9 | 978 | 9 | HSA323696 | AJ323696 Homo sapi |
| 11 | 675.6 | 56.7 | 1247 | 10 | AF366401 | AF366401 Mus muscu |
| 12 | 667.4 | 56.0 | 5869 | 6 | AX537459 | AX537459 Sequence |
| 13 | 640.8 | 53.8 | 1165 | 10 | BC016077 | BC016077 Mus muscu |
| 14 | 637.2 | 53.5 | 696 | 6 | BD124736 | BD124736 Primer fo |
| 15 | 637.2 | 53.5 | 696 | 6 | BD126760 | BD126760 Primer fo |
| 16 | 588 | 49.3 | 145347 | 2 | AC073821 | AC073821 Mus muscu |
| 17 | 573.4 | 48.1 | 877 | 10 | AF218291 | AF218291 Mus muscu |
| 18 | 569.6 | 47.8 | 201377 | 2 | AC073767 | AC073767 Mus muscu |
| 19 | 553.6 | 46.4 | 257945 | 2 | AC112313 | AC112313 Rattus no |
| 20 | 553.6 | 46.4 | 290929 | 2 | AC118914 | AC118914 Rattus no |
| 21 | 532.4 | 44.7 | 164652 | 2 | AC021625 | AC021625 Homo sapi |
| 22 | 501 | 42.0 | 701 | 10 | AF366400 | AF366400 Mus muscu |
| 23 | 433.8 | 36.4 | 450 | 6 | BD058236 | BD058236 Secreted |
| 24 | 419.8 | 35.2 | 745 | 9 | HSA325585 | AJ325585 Homo sapi |
| 25 | 391.6 | 32.9 | 476 | 6 | BD109590 | BD109590 EST and e |
| 26 | 330 | 27.7 | 256418 | 2 | AC115309 | AC115309 Rattus no |
| 27 | 293.4 | 24.6 | 301 | 6 | BD076276 | BD076276 5' EST of |
| 28 | 237.6 | 19.9 | 536 | 6 | AX283999 | AX283999 Sequence |
| 29 | 227.6 | 19.1 | 68118 | 2 | AC011113 | AC011113 Homo sapi |
| 30 | 172.8 | 14.5 | 986 | 9 | HSA323697 | AJ323697 Homo sapi |
| 31 | 69.6 | 5.8 | 125020 | 9 | AF429315 | AF429315 Homo sapi |
| 32 | 67.2 | 5.5 | 173048 | 5 | AL772154 | AL772154 Zebrafish |
| 33 | 66 | 5.5 | 173048 | 5 | AL772154 | AL772154 Zebrafish |
| 34 | 63.4 | 5.3 | 127051 | 2 | AC114895 | AC114895 Tetradon |
| 35 | 59.8 | 5.0 | 134408 | 2 | AC139628 | AC139628 Takifugu |
| 36 | 58.6 | 4.9 | 2000 | 6 | AX655393 | AX655393 Sequence |
| 37 | 50.4 | 4.2 | 314100 | 1 | SC039106 | AL939106 Streptomy |
| 38 | 47.4 | 4.0 | 2000 | 6 | AX655393 | AX655393 Sequence |
| 39 | 47 | 3.9 | 1885 | 9 | AF514298 | AF514298 Homo sapi |
| 40 | 47 | 3.9 | 1969 | 9 | AF093445 | AF093445 Homo sapi |
| 41 | 47 | 3.9 | 138999 | 9 | AC090426 | AC090426 Homo sapi |
| 42 | 47 | 3.9 | 158766 | 2 | AC015851 | AC015851 Homo sapi |
| 43 | 47 | 3.9 | 166234 | 9 | AC018629 | AC018629 Homo sapi |
| 44 | 47 | 3.9 | 173440 | 2 | AC080112 | AC080112 Homo sapi |
| 45 | 47 | 3.9 | 188574 | 2 | AC131063 | AC131063 Homo sapi |

ALIGNMENTS

RESULT 1
BD127703
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 2113)
REFERENCE
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
BD127703 2113 bp DNA linear PAT 18-SEP-2002
Primer for synthesizing full-length cDNA and use thereof.
BD127703
BD127703.1 GI:23222648
JP 2002017375-A/3134.
Homo sapiens (human)
Homo sapiens

Pred. No. is the number of results predicted by chance to have a

TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017375-A 3134 22-Jan-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/3134
PD 22-Jan-2002
PF 07-Jul-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOgai, KOJI HAYASHI, SHIZUKO
PI YORI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHIMICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA

PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key

FEATURES
source 1. 2113
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 453 a 561 c 578 g 521 t
ORIGIN

Query Match 99.3%; Score 1183.4; DB 6; Length 2113;
Best Local Similarity 99.9%; Pred. No. 2.3e-251;
Matches 1184; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GTGACCCGGAGTCAACAACCTGCTCGCGAGACCCAGCGGGGTTCCGGAGAGACCCAA 60
21 GTGACCCGGAGTCAACAACCTGCTCGCGAGACCCAGCGGGGTTCCGGAGAGACCCAA 80
61 CAAGGATGCTGCGCCCGCGCTTCTGATTTGGTGGTGGGACCTGCTGCTGCTGCTGCTG 120
81 CAAGGATGCTGCGCCCGCGCTTCTGATTTGGTGGTGGGACCTGCTGCTGCTGCTGCTG 140
121 TTGCGCGCTAGTGAACAAGATGCTGAGCAAGGGTCTGAAGCGGAAGCGGAGAGAGAGA 180
141 TTGCGCGCTAGTGAACAAGATGCTGAGCAAGGGTCTGAAGCGGAAGCGGAGAGAGAGA 200
181 GGAGAGAGAACTCTGCGAGTGCAGTCTGCTGGTGGTGAATCTCTGGCCACGACGGTGGC 240
201 GGAGAGAGAACTCTGCGAGTGCAGTCTGCTGGTGGTGAATCTCTGGCCACGACGGTGGC 260
241 ACAGGACACCCCGCGCGCTGAGCTCCCTTTGACCTCTGACCTCAAGCTCAACCTCA 300
261 ACAGGACACCCCGCGCGCTGAGCTCCCTTTGACCTCTGACCTCAAGCTCAACCTCA 320
301 CCACAGCTGACGAGAGTGAAGCGGACCTGCGGACCTGGTGGTGGTGGTGGTGGTGGT 360
321 CCACAGCTGACGAGAGTGAAGCGGACCTGCGGACCTGGTGGTGGTGGTGGTGGTGGT 380
361 GCGGCGCATCCAGCGCTGATGAGCAACCGCGGCGCTGCTGCGACCTGCTGCTGCTGCTG 420
381 GCGGCGCATCCAGCGCTGATGAGCAACCGCGGCGCTGCTGCGACCTGCTGCTGCTGCTG 440
421 TGCAGACCCCGCGCGCTGAGCAACTTACTGGCAAGCTCGAGACGCTGCTGCTGCTGCTG 480
441 TGCAGACCCCGCGCGCTGAGCAACTTACTGGCAAGCTCGAGACGCTGCTGCTGCTGCTG 500
481 CATGGCCAGCTCTCTGAGAGACTCAGACACATTTAGAGGAGCTGAGTCAAGCTCCCAAC 540
501 CATGGCCAGCTCTCTGAGAGACTCAGACACATTTAGAGGAGCTGAGTCAAGCTCCCAAC 560
541 CTTGGAGAGAGAGGAGGAGCCAGGACCGTGAATCGGGGAGACAGCGCCACCTGGGTGC 600
561 CTTGGAGAGAGAGGAGGAGCCAGGACCGTGAATCGGGGAGACAGCGCCACCTGGGTGC 620
601 CTTGAGACTGCTGGGCGGACGCACTGGCTGCTACTGAGACATGGCGCTTGAAGGCGCTGT 660

Db 621 CTTGGACCTGCTGGGCGGACGCACTGCTGCTACTGAGCATGGGCTTGAAGGCGCTGT 680
Qy 661 TGAGGATATTGACACCTCTATGATATGACATGAATGTTGGGACACGCTCTGAAGGCT 720
Db 661 TGAGGATATTGACACCTCTATGATATGACATGAATGTTGGGACACGCTCTGAAGGCT 740
Qy 721 CAACACGAGCCCTGAGATGGGCGCGGAGAGAGAGAGCTCCGAGCTGACGAGGCGGA 780
Db 741 CAACACGAGCCCTGAGATGGGCGCGGAGAGAGAGAGCTCCGAGCTGACGAGGCGGA 800
Qy 781 ATTGACATCTCAATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 801 ATTGACATCTCAATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 860
Qy 841 AGGCGCTGAGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 861 AGGCGCTGAGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 920
Qy 901 GACCACTGCTCTGCAAAAAGACACAGCTGGCTTCCCTAGTACAGAGAAACAGGCTGGGC 960
Db 921 GACCACTGCTCTGCAAAAAGACACAGCTGGCTTCCCTAGTACAGAGAAACAGGCTGGGC 980
Qy 961 CACTTGGAGAGACAGATCTAGTCTGCTGGGCACTTACATCCGCTGCTGCTGCTGCTGCTG 1020
Db 981 CACTTGGAGAGACAGATCTAGTCTGCTGGGCACTTACATCCGCTGCTGCTGCTGCTGCTG 1040
Qy 1021 CTGGCAGGAGGAGCTGGAATTTACCCCTAGTATGATGATGACAGAGGCTGCTGGGACT 1080
Db 1041 CTGGCAGGAGGAGCTGGAATTTACCCCTAGTATGATGATGACAGAGGCTGCTGGGACT 1100
Qy 1081 GAATTCCTGCGCCCTGGGCTATGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 1101 GAATTCCTGCGCCCTGGGCTATGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1160
Qy 1141 CCAATGACATTTTCAATTAAGCAAGCTCTGGGAATCTCAAA 1185
Db 1161 CCAATGACATTTTCAATTAAGCAAGCTCTGGGAATCTCAAA 1205

RESULT 2
AK074652
LOCUS
DEFINITION Homo sapiens cDNA FLJ90171 fis, clone MAMMA1000403, highly similar
to Homo sapiens CD4-binding protein p34SE11 (SE11) mRNA.
ACCESSION AK074652
VERSION AK074652.1 GI:22760232
KEYWORDS oligo cloning; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 Isogai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T.,
Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S.,
Kawai, H., Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y.,
Kojima, S., Negami, K., Masuno, Y., Ono, T., Okano, K., Yoshikawa, Y.,
Aotsuka, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and
Ninomiya, K.

TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2113)
AUTHORS Isogai, T. and Otsuki, T.
DIRECT SUBMISSION
JOURNAL Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yano, Risazazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction:
Institute of Medical Science, University of Tokyo, Laboratory of
Genome Structure, Human Genome Center; cDNA 5'-8' 3'-end one pass
sequencing and clone selection: Helix Research Institute (supported
by Japan Key Technology Center etc.).

```

FEATURES
  source
    Location/Qualifiers
      1..2113
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="MAMMA1000403"
      /tissue_type="mammary gland"
      /clone_lib="MAMMA1"
      /note="cloning vector: pME18SFL3"

BASE COUNT      453 a      561 c      578 g      521 t

ORIGIN
Query Match      99.3%; Score 1183.4; DB 9; Length 2113;
Best Local Similarity 99.9%; Pred. No. 2.3e-251;
Matches 1184; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GTGAGCGGAGTAGAAGTGCCTCTGCGAGCCGAGCGGGGGTTCCGGAGGACACCA 60
DB      21 GTAGCGCGAGTAGAAGTGCCTCTGCGAGCCGAGCGGGGGTTCCGGAGGACACCA 80
QY      61 CAAGCATGCTGCGCGCGCGCTTCTGATTTGGTTGGGTGAGTACTCTTCTGTTCTGA 120
DB      81 CAAGCATGCTGCGCGCGCGCTTCTGATTTGGTTGGGTGAGTACTCTTCTGTTCTGA 140
QY      121 TTGGCCGCTAGTAGACAAGTCTGACGAAGGCTCTGAAGCGGAACGGAGAGAGA 180
DB      141 TTGGCCGCTAGTAGACAAGTCTGACGAAGGCTCTGAAGCGGAACGGAGAGAGA 200
QY      181 GGAGAGGAACCTCTGGAGTGCATGCTCGTGGTGGCTAGATCTTGGGCGACGACGGTGC 240
DB      201 GGAGAGGAACCTCTGGAGTGCATGCTCGTGGTGGCTAGATCTTGGGCGACGACGGTGC 260
QY      241 ACAGGACCCCGCGCGCTGAGCTCTAGCTCCCTCTTTGAACCTCTCAGTGCCTCAAGCTCA 300
DB      261 ACAGGACCCCGCGCGCTGAGCTCTAGCTCCCTCTTTGAACCTCTCAGTGCCTCAAGCTCA 320
QY      301 CCACAGCTCTGACAGAGTAGAGCCGACCTGCGGCACTGTGTGCTGTGTAACACTCT 360
DB      321 CCACAGCTCTGACAGAGTAGAGCCGACCTGCGGCACTGTGTGCTGTGTAACACTCT 380
QY      361 GCGGCGCATCGAGGCTCCATGAGCAGCCGCGGCTGCGCCACCTGTGCTTACGCCAC 420
DB      381 GCGGCGCATCGAGGCTCCATGAGCAGCCGCGGCTGCGCCACCTGTGCTTACGCCAC 440
QY      421 TGCAGCCCGCCAGTGTGCTGACAACTTACTGCGACAGCTGAGAGCTGCCCTTACAGCTTC 480
DB      441 TGCAGCCCGCCAGTGTGCTGACAACTTACTGCGACAGCTGAGAGCTGCCCTTACAGCTTC 500
QY      481 CATGCGCAGCTCTCTGAGAGACTCTGACCAATTTGAGGGCTGTGAGCTCCCAACC 540
DB      501 CATGCGCAGCTCTCTGAGAGACTCTGACCAATTTGAGGGCTGTGAGCTCCCAACC 560
QY      541 CTGGGAGAGAGAGGGGCGACAGAGCGGTGATCGGGGGGAGAGCGCCCAAGCTGGGTGC 600
DB      561 CTGGGAGAGAGAGGGGCGACAGAGCGGTGATCGGGGGGAGAGCGCCCAAGCTGGGTGC 620
QY      601 CTGGGAGAGAGAGGGGCGACAGAGCTGCTCTACTGAGAGATGGGCTTGAAGGGCTGTT 660
DB      621 CTGGGAGAGAGAGGGGCGACAGAGCTGCTCTACTGAGAGATGGGCTTGAAGGGCTGTT 680
QY      661 TGAAGATATTTGACACTCTTATGTATGACAAATGAACTTTGGGACCAAGCCTCTGAGGGCCT 720
DB      681 TGAAGATATTTGACACTCTTATGTATGACAAATGAACTTTGGGACCAAGCCTCTGAGGGCCT 740
QY      721 CAAACAGAGCGCTGAGAGATGGGCGGGCAAGAGGAAGCTCGGAGCTGGAGAGAGCCCA 780
DB      741 CAAACAGAGCGCTGAGAGATGGGCGGGCAAGAGGAAGCTCGGAGCTGGAGAGAGCCCA 800
QY      781 ATTGAGTACTATGATGTGCTGTGGGCAACAGGAGACTGAGAGCGAGCGCGGGGCG 840
DB      801 ATTGAGTACTATGATGTGCTGTGGGCAACAGGAGACTGAGAGCGAGCGCGGGGCG 860
QY      841 AGGGCGCTGAGCCCTGCTGCTGGAGATGTTGTCTGTTATCTGAACCTGTGCTGCTG 900

```

```

DB      861 AGGGCGCTGAGCCCTGTGCTGGAATGTTGTCTGATCTGAACAGAGCTGCTGGCTG 920
QY      901 GACCACTGTCCTCGAAAAGACACACAGCTGCTCCCTAGTACAGATAACAGGGCTGGCG 960
DB      921 GACCACTGTCCTCGAAAAGACACACAGCTGCTCCCTAGTACAGATAACAGGGCTGGCG 980
QY      961 CACTTTGAGAGACAGAAATCTAGTCTGGGCAACTTACATCCGCTCCTGCTGTCAAGG 1020
DB      981 CACTTTGAGAGACAGAAATCTAGTCTGGGCAACTTACATCCGCTCCTGCTGTCAAGG 1040
QY      1021 CTGGCAGGGGGAGCCTGGAATTACCCCTAGTATGATGAATGACAGGCTGTGAGGACT 1080
DB      1041 CTGGCAGGGGGAGCCTGGAATTACCCCTAGTATGATGAATGACAGGCTGTGAGGACT 1100
QY      1081 GAATTCCTGGGCGCTGGGGCATAGCTTGGGGCTGCTCTCTGTGATPAGGGAAGAGACC 1140
DB      1101 GAATTCCTGGGCGCTGGGGCATAGCTTGGGGCTGCTCTCTGTGATPAGGGAAGAGACC 1160
QY      1141 CCAATCAGATTTTTCAAATTAAAGCCAGTCTGGGAATCTCAAA 1185
DB      1161 CCAATCAGATTTTTCAAATTAAAGCCAGTCTGGGAATCTCAAA 1205

```

```

RESULT 3
BC002670
LOCUS
DEFINITION
  Homo sapiens, CDK4-binding protein p34SE11, clone MGC:3465
ACCESSION
  BC002670
VERSION
  BC002670.1 GI:12803668
KEYWORDS
  MGC.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
  1 (bases 1 to 1179)
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  Strausberg, R.
  Direct Submission
  Submitted (05-FEB-2001) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
  USA

```

```

REMARK
  NIH-MGC Project URL: http://mgc.ncl.nih.gov
  Contact: MGC help desk
  Email: cgabs-remail.nih.gov
  Tissue Procurement: ATCC

```

```

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
  Gaithersburg, Maryland;
  Web site: http://www.nisc.nih.gov/

```

```

Contact: nisc_mgc@ncl.nih.gov
  Shevchenko, Y., Wehlerby, K.D., Beckstrom-Sternberg, S.M.,
  Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
  Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
  Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
  McDowell, J., Pearson, R., Snyder, B., Stantirpop, S., Thomas, P.J.,
  Tlionson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
  Zhang, L.-H. and Green, E.D.

```

```

Clone distribution: MGC clone distribution information can be found
  through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
  Series: IRAL Plate: 12 Row: i Column: 23
  This clone was selected for full length sequencing because it
  passed the following selection criteria: matched mRNA gi: 7019524.

```

```

Location/Qualifiers
  1..1179

```

```

FEATURES
  source
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /db_xref="taxon:9606"

```

```

/clone="MGC:3465 IMAGE:3613213"
/clone_type="uterus", endometrium adenocarcinoma"
/clone_id="NIH_MGC_44"
/lab_host="DH10B-R"
/notes="Vector: pORB7"
118..828
/codon_start=1
/product="CDK4-binding protein p34SEI1"
/protein_id="AAH02670.1"
/db_xref="GI:12803669"
/translation="MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAPPAVASS
LFDLSVTLKHSLSQSEPDRLRHLYLVTLRIQASMAPAALPVPSPAPASVDN
LIASDPAALSASMSLLEDLISHIEGLSQAAPPLADBPGRSISGAAPSLGDLILGP
ATGCLLDGLEGLEPDIDTYSMDNELMIPASEGLKPGEDPGRKEAELEADLEIDL
MDVAVGTQALRRPPGR"
BASE COUNT      243 a      343 c      368 g      225 t
ORIGIN
Query Match      97.3% Score 1159.4; DB 9; Length 1179;
Best Local Similarity 99.9%; Pred. No. 5e-246;
Matches 1160; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 32 CCAGGCGGGGTTTCGGAGACACCAACAAGCGATGTCGCCGCCGTTTCTGAT 91
DB 10 CCAGGCGGGGTTTCGGAGACACCAACAAGCGATGTCGCCGCCGTTTCTGAT 69
OY 92 GATTGGGTGGCTACCTCTCTGTTCTGATTGGCCGCTAGTAGAAGATGCTGACAG 151
DB 70 GATTGGGTGGCTACCTCTCTGTTCTGATTGGCCGCTAGTAGAAGATGCTGACAG 129
OY 152 GGTGGAAGCGGAACGGAGAGAGAGAGGAAGAAAGCTCTGGCAGTGCATCCCTGG 211
DB 130 GGTGGAAGCGGAACGGAGAGAGAGAGGAAGAAAGCTCTGGCAGTGCATCCCTGG 189
OY 212 TGGCTAGATCTGGCCACAGCGGTGGACAGGACCCCGCGCTGCTAGCTTC 271
DB 190 TGGCTAGATCTGGCCACAGCGGTGGACAGGACCCCGCGCTGCTAGCTTC 249
OY 272 CTCTTTGACCTCTCAGTCTCAGTCTCAGACACAGCTGACAGAGTGAAGCCGAGCT 331
DB 250 CTCTTTGACCTCTCAGTCTCAGTCTCAGACACAGCTGACAGAGTGAAGCCGAGCT 309
OY 332 CGGACCGGTTGCTGGTGTGTAACACTGTGGGGGACATCCAGGGCTCCATGGACACCCGG 391
DB 310 CGGACCGGTTGCTGGTGTGTAACACTGTGGGGGACATCCAGGGCTCCATGGACACCCGG 369
OY 392 GCTGCCCTGACCACTGTGCTAGCCACACTGAGCCGCCAGTGTGGCTGACAACTTACTG 451
DB 370 GCTGCCCTGACCACTGTGCTAGCCACACTGAGCCGCCAGTGTGGCTGACAACTTACTG 429
OY 452 GCAAGCTGGACGCTGCCCTTTCAAGCTTCATGGCCAGCTCTCTGAGAGACCTCAGCCAG 511
DB 430 GCAAGCTGGACGCTGCCCTTTCAAGCTTCATGGCCAGCTCTCTGAGAGACCTCAGCCAG 489
OY 512 ATTGAGGCGCTGAGTCAAGGCTCCCAACCTTTGGAGAGAGAGGGGCCACAGGCCGTAGC 571
DB 490 ATTGAGGCGCTGAGTCAAGGCTCCCAACCTTTGGAGAGAGAGGGGCCACAGGCCGTAGC 549
OY 572 ATCGGGGAGAGCGGCCACAGCTGGGTGCTTGAAGCTGTTGGGCCACAGCCACTGGCTGT 631
DB 550 ATCGGGGAGAGCGGCCACAGCTGGGTGCTTGAAGCTGTTGGGCCACAGCCACTGGCTGT 609
OY 632 CTACTGGACGATGGGCTTGAAGGCTGTTTGAAGATATATGACACTCTATATGACAAAT 691
DB 610 CTACTGGACGATGGGCTTGAAGGCTGTTTGAAGATATATGACACTCTATATGACAAAT 669
OY 692 GAATTTGGACACAGCTCTGAGGGCTCAACACAGGACCTGAGAGATGGGCGGGAG 751
DB 670 GAATTTGGACACAGCTCTGAGGGCTCAACACAGGACCTGAGAGATGGGCGGGAG 729
OY 752 GAGGAAGCTCCGAGCTGGAGAGGCGGAATTTGACTACCTCATGATGTCTGATGGGC 811
DB 730 GAGGAAGCTCCGAGCTGGAGAGGCGGAATTTGACTACCTCATGATGTCTGATGGGC 789

```

```

OY 812 ACACAGCAGCTGGAGCGACCCGCCGGCCAGAGCCCTAGCCCTCTGCTGGAATGGTTG 871
DB 790 ACACAGCAGCTGGAGCGACCCGCCGGCCAGAGCCCTAGCCCTCTGCTGGAATGGTTG 849
OY 872 TCTGTATCTGAGTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 931
DB 850 TCTGTATCTGAGTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 909
OY 932 TTCCCTAGTACAGAGACAGGAGCTTGGGCACTTTGGAGAGACAAATCTAGTCTGGGC 991
DB 910 TTCCCTAGTACAGAGACAGGAGCTTGGGCACTTTGGAGAGACAAATCTAGTCTGGGC 969
OY 992 AACTTCACATCCGCTCCTCTCTCTGAGGCTGAGAGGGGAGGAGCTGGAATTAACCCCTAG 1051
DB 970 AACTTCACATCCGCTCCTCTCTCTGAGGCTGAGAGGGGAGGAGCTGGAATTAACCCCTAG 1029
OY 1052 TGATGGAATGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1111
DB 1030 TGATGGAATGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1089
OY 1112 CTGTTCTCTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1171
DB 1090 CTGTTCTCTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1149
OY 1172 TGGGAATCTCAAAAAA 1192
DB 1150 TGGGAATCTCAAAAAA 1170
RESULT 4
AF117959
LOCUS      Homo sapiens CDK4-binding protein p34SEI1 (SEI1) mRNA, complete cds.
DEFINITION
VERSION    AF117959.1 GI:6434875
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1218)
AUTHORS   Sugimoto,M., Nakamura,T., Ohtani,N., Hampson,L., Hampson,I.N.,
           Shimamoto,A., Futuchi,Y., Okumura,K., Niwa,S., Taya,Y. and Hara,E.
           Regulation of CDK4 activity by a novel CDK4-binding protein,
           p34(SEI-1)
JOURNAL    Genes Dev. 13 (22), 3027-3033 (1999)
MEDLINE    20047903
PUBMED     10580009
REFERENCE  2 (bases 1 to 1218)
AUTHORS   Hara,E. and Sugimoto,M.
JOURNAL    Direct Submission
TITLE      Submitted (05-JAN-1999) Cell Cycle Laboratory, Paterson Institute
           for Cancer Research, Wilmslow Road, Manchester M20 4BX, UK
FEATURES
SOURCE
1..1218
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.1-q13.2"
1..1218
/gene="SEI1"
131..841
/gene="SEI1"
/codon_start=1
/product="CDK4-binding protein p34SEI1"
/protein_id="AAH08349.1"
/db_xref="GI:6434876"
/translation="MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAPPAVASS
LFDLSVTLKHSLSQSEPDRLRHLYLVTLRIQASMAPAALPVPSPAPASVDN
LIASDPAALSASMSLLEDLISHIEGLSQAAPPLADBPGRSISGAAPSLGDLILGP

```



```

Db 122 AGCAAGATGCTGAGCAAGGGTCTGTAAGGGGAACGGGAGAGAGAGAGAGAAAGAACT 181
OY 194 CTGGCAGTGCAGCTCCTGGTGGCTAGATCCTGGCCACGAGCGGTGGACAGAGCCACCCCG 253
Db 182 CTGGCAGTGCAGCTCCTGGTGGCTAGATCCTGGCCACGAGCGGTGGACAGAGCCACCCCG 241
OY 254 GCGCGGGCTCTAGCTCCCTCTTTTACCTCTCAGTGCCTCAAGCTCCACACAGCTCTGAG 313
Db 242 GCGCGGGCTCTAGCTCCCTCTTTTACCTCTCAGTGCCTCAAGCTCCACACAGCTCTGAG 301
OY 314 CAGAGTGCAGCGGACCTCGGACCTGGTGCCTGGTGAACACTCTGGGGGCGATCCAG 373
Db 302 VRRAGTBRAGCGGAGCTCGGACCTGGTGCCTGGTGAACACTCTGGGGGCGATCCAG 361
OY 374 GCGTCCATGAGCAACCGCGGCTGCCCTGCACCTGTGCTAGCCACCTGCAGCCGCCAGT 433
Db 362 GCGTCCATGAGCAACCGCGGCTGCCCTGCACCTGTGCTAGCCACCTGCAGCCGCCAGT 421
OY 434 GTGGCTGACAACTTACTGGCAAGCTCGGACCTGGCTTTTCAAGCTCCATGGCCAGCTC 493
Db 422 GTGGCTGACAACTTACTGGCAAGCTCGGACCTGGCTTTTCAAGCTCCATGGCCAGCTC 481
OY 494 CTGGAGAGCTCAGCCACATTTAGAGGGCTGAGTCAAGGCTCCCAACCTTTGGCAGAGAG 553
Db 482 CTGGAGAGCTCAGCCACATTTAGAGGGCTGAGTCAAGGCTCCCAACCTTTGGCAGAGAG 541
OY 554 GGGCCACAGCGCCGTAGCATCGGGGAGAGCGCCAGCCAGCTGGTGCCTTGGACCTGCTG 613
Db 542 GGGCCACAGCGCCGTAGCATCGGGGAGAGCGCCAGCCAGCTGGTGCCTTGGACCTGCTG 601
OY 614 GGGCCACAGCGCTGTACTGAGCATGGGCTTGAGGGCTTTTGAAGATTTTAC 673
Db 602 GGGCCACAGCGCTGTACTGAGCATGGGCTTGAGGGCTTTTGAAGATTTTAC 661
OY 674 ACCTTATGATGACAAATGAACTTTGGGACACAGCTCTGAGGGCTTCAAAACAGGCTT 733
Db 662 ACCTTATGATGACAAATGAACTTTGGGACACAGCTCTGAGGGCTTCAAAACAGGCTT 721
OY 734 GAGGATGGGGCGGGCAGAGAGAACTCCGAGCTGAGAGAGCCGAATTTGACTACTCTC 793
Db 722 GAGGATGGGGCGGGCAGAGAGAACTCCGAGCTGAGAGAGCCGAATTTGACTACTCTC 781
OY 794 ATGATGCTGTGGGGCAGACAGGCACTGAGAGAGCCGGGGCCAGGGCGCTGAGAGC 853
Db 782 ATGATGCTGTGGGGCAGACAGGCACTGAGAGAGCCGGGGCCAGGGCGCTGAGAGC 841
OY 854 CTGCTGCTGAAATGATGCTGTGATCTGAATCTGAGCTGCTGAGTGGACCACTGCTCT 913
Db 842 CTGCTGCTGAAATGATGCTGTGATCTGAATCTGAGCTGCTGAGTGGACCACTGCTCT 901
OY 914 CGAAAAGACACAGCTGCTCCCTAGTACAGAGAAACAGGCTTGGGCCACTTTGGAGAGA 973
Db 902 CGAAAAGACACAGCTGCTCCCTAGTACAGAGAAACAGGCTTGGGCCACTTTGGAGAGA 961
OY 974 CAGAAATCAGTCCCTGGGCAACTTTCACATCCCTCTCTGCTCAAGGGCTGGCAGAGGGAG 1033
Db 962 CAGAAATCAGTCCCTGGGCAACTTTCACATCCCTCTCTGCTCAAGGGCTGGCAGAGGGAG 1021
OY 1034 CCTGGAATTTACCCCTAGTATGATGAAATGACAGGCTTGGGGACTGAATTTCCCTGGCC 1093
Db 1022 CCTGGAATTTACCCCTAGTATGATGAAATGACAGGCTTGGGGACTGAATTTCCCTGGCC 1081
OY 1094 CTGGGGCTCATAGCTTGGGGCTTCTCTCTGATATACGGGAGAGAGCCCAATCAGATTTT 1153
Db 1082 CTGGGGCTCATAGCTTGGGGCTTCTCTCTGATATACGGGAGAGAG-CCCAATCAGATTTT 1140
OY 1154 TCAAAATTAAGCCAGTCTCTGGGAAATCTCAAAAAAAA 1192
Db 1141 TCAAAATTAAGCCAGTCTCTGGGAAATCTCAAAAAAAA 1179

```

RESULT 6
AF366402

```

LOCUS AF366402 1178 bp mRNA linear PRI 13-MAY-2001
DEFINITION Homo sapiens TRIP-Brl mRNA, complete cds.
ACCESSION AF366402
VERSION AF366402.1 GI:14029833
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1178)
Hsu,S.-I., Yang,C.-M., Slim,K.G., Hentschel,D.M., O'Leary,E. and
Bonventre,J.V.
TRIP-Brl: a novel family of PHD zinc finger- and
bromodomain-interacting proteins that regulate the transcriptional
activity of E2F-1/DP-1
EMBO J. 20 (9), 2273-2285 (2001)
JOURNAL MEDLINE 21231173
PUBMED 11331592
REFERENCE 2 (bases 1 to 1178)
Hsu,S.-I. and Bonventre,J.V.
Direct Submision
JOURNAL Submitted (29-MAR-2001) Dept. of Medicine, National University
Hospital, 5 Lower Kent Ridge Road 119074, Singapore
FEATURES
source
1..1178
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="dbEST:W52056"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"
/feature="19q13"
/function="acts at E2F-responsive promoters to integrate
signals provided by PHD- and/or bromodomain-containing
transcription factors"
/function="a CDK4-binding protein that renders the
activity of cyclin D/CDK4 resistant to the inhibitory
effects of p16 (INK4a)"
/note="hTRIP-Brl; transcriptional regulator interacting
with the PHD-bromodomain; hTRIP-Brl(SEI-1)"
/codon_start=1
/product="TRIP-Brl"
/protein_id="AAK52831.1"
/db_xref="GI:14029834"
/translation="MLSGLKRRKEEKEEPLAVDSMWLDPGHAAYAAQAPPAVASS
LFDLISVLKLSHLSLOOSEPDRLVLYVNTLRIOASMAAPALPPSPAPSAVAN
LLASDDALSASMSLIEDLSHIEGLSOAPDLPDPPGSGAAPSIGALDILAP
ATGCLLDGLEGLIEDLIDTSMYDNLMAAPASEGLKPPEDGPKKEAPRLDENLDTL
MDVTVGQALERPPGPRG"
BASE COUNT 245 a 343 c 365 g 225 t
ORIGIN
Query Match 94.8%; Score 1130.4; DB 9; Length 1178;
Best Local Similarity 99.1%; Pred. No. 1.3e-229;
Matches 1158; Conservative 0; Mismatches 6; Indels 4; Gaps 2;
OY 26 CGCAGACCAGGCGGGGTTTCCGGAGACACCAACAGAGATCGCCGCGGCTTTC 85
Db 1 CGCAGACCAGGCGAG--TTTCCGAGAGGACACCAACAGATCTCGCGCGGCTTTC 57
OY 86 CTGATTTGGTGTGGGTGCTACCTCTTCTGATTTGGCCGCTAGTAGACAGATCTG 145
Db 58 CTGATTTGGTGTGGGTGCTACCTCTTCTGATTTGGCCGCTAGTAGACAGATCTG 117
OY 146 AGCAAGGCTCGAAGCGGAACGGAGAGAGAGAGAGAGAAAGCACTTCGCGCTGAC 205
Db 118 AGCAAGGCTCGAAGCGGAACGGAGAGAGAGAGAGAGAAAGCACTTCGCGCTGAC 177
OY 206 TCTGTGTGCTAGATCTCTGGCCAGCAGCGGTGGCAGACAGCCCGCGGCTGCTCT 265
Db 178 TCTGTGTGCTAGATCTCTGGCCAGCAGCGGTGGCAGACAGCCCGCGGCTGCTCT 237
OY 266 AGCTTCCTCTTTGACCTCTCAGTCTCAGCTCCACCAACAGCTGACAGAGTGAAGCG 325

```

```

Db      238 AGCTCCCTCTTGTACCTCTCAGCTCAGCTCAGCTCCACACAGCTCGAGAGAGATGAGCCG 297
QY      326 AGCTTCGGGCGACCTGGTGTGCTGTGAACACTCTGCGGCGATCCAGGCGTCCAGGACA 385
Db      298 GACCTCGGGACACTGGTGTGCTGTGAACACTCTGCGGCGATCCAGGCGTCCAGGACA 357
QY      386 CCGCGGGCTCCCTGCGACCTGTGCTGTGAACACTCTGCGGCGATCCAGGCGTCCAGGACA 445
Db      358 CCGCGGGCTCCCTGCGACCTGTGCTGTGAACACTCTGCGGCGATCCAGGCGTCCAGGACA 417
QY      446 TTACTGGAAGCTCGGAGCGCTGCTGTGAGCTTCATGCGCAGCTTCCTGAGAGCGCTC 505
Db      418 TTACTGGAAGCTCGGAGCGCTGCTGTGAGCTTCATGCGCAGCTTCCTGAGAGCGCTC 477
QY      506 AGCACAATTTAGAGGCGCTGAGTACAGCTCCCAACCTTTGGACAGAGAGGCGCCACGAGC 565
Db      478 AGCACAATTTAGAGGCGCTGAGTACAGCTCCCAACCTTTGGACAGAGAGGCGCCACGAGC 537
QY      566 CGTAGCATTTGGGGGAGCAGCCCGCAGCTGGGTGCTGCTGAGCTGCTGGCCGAGCAGCT 625
Db      538 CGTAGCATTTGGGGGAGCAGCCCGCAGCTGGGTGCTGCTGAGCTGCTGGCCGAGCAGCT 597
QY      626 GGCTGCTACTGAGCATGGGCTTGAAGGCGCTTTGAGGATATTGACACCTTATGTAT 685
Db      598 GGCTGCTACTGAGCATGGGCTTGAAGGCGCTTTGAGGATATTGACACCTTATGTAT 657
QY      686 GACAATGAATTTGGGCGACAGCTCTGAGAGGCGCTCAACAGCCCTTGAAGTGGGCGG 745
Db      658 GACAATGAATTTGGGCGACAGCTCTGAGAGGCGCTCAACAGCCCTTGAAGTGGGCGG 717
QY      746 GCGAAGAGAGAACTCCGAGCTGAGAGAGCGCGAATTTGAGTACCTCATGTGATGCTG 805
Db      718 GCGAAGAGAGAACTCCGAGCTGAGAGAGCGCGAATTTGAGTACCTCATGTGATGCTG 777
QY      806 GTGGGCGACAGGCACTGAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 865
Db      778 GTGGGCGACAGGCACTGAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 837
QY      866 TGGTGTCTGTGATCTGAACTGAGCTGCTGCTGAGCCCACTGCTCTCGAAAGACACA 925
Db      838 TGGTGTCTGTGATCTGAACTGAGCTGCTGCTGAGCCCACTGCTCTCGAAAGACACA 897
QY      926 GCTGGCTTCCCTAGTACAGAGAGAGCGCTTGGGCGCTTTGGAGAGAGACAGATCTAGTC 985
Db      898 GCTGGCTTCCCTAGTACAGAGAGAGCGCTTGGGCGCTTTGGAGAGAGACAGATCTAGTC 957
QY      986 CTGGGCAACTTCACATCCGCTCCCTGCTCAAGGCGTGGCA-CGGGAGACCTGGAAATTAC 1044
Db      958 CTGGGCAACTTCACATCCGCTCCCTGCTCAAGGCGTGGCAAGGCGGAGCCTGGAAATTAC 1017
QY      1045 CCCCTAGTATGAAATGACAGAGGCTGTGGGAGTGAATTCCTGCGGCGGCGGCGGCA 1104
Db      1018 CCCCTAGTATGAAATGACAGAGGCTGTGGGAGTGAATTCCTGCGGCGGCGGCGGCA 1077
QY      1105 GCTTGGGCTGTCTCTCTCTGTGATACGGAAGAGACCCCAATGATTTTCAAAATTAAAG 1164
Db      1078 GCTTGGGCTGTCTCTCTCTGTGATACGGAAGAGACCCCAATGATTTTCAAAATTAAAG 1137
QY      1165 CCAGTCCCTGGAAATCTCAAAAAAAA 1192
Db      1138 CCAGTCCCTGGAAATCTCAAAAAAAA 1165

```

```

RESULT 7
LOCUS   AY130860          5917 bp    DNA       linear    PRI 22-JUL-2002
DEFINITION Homo sapiens SCDK4-binding protein p34SEI1 (SEI1) gene, complete
ACCESSION AY130860
VERSION   AY130860.1   GI:21914609
KEYWORDS
SOURCE    Homo sapiens (human)

```

```

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5917)
Rieder M.J., Braun A.C., Montoya M.A., Chung M.-W., Nguyen C.P.,
Nguyen D.A., Livingston R.J., Foell C.L., Robertson P.D.,
Schackwitz W.S., Sherman J.K., Wiltak L.A. and Nickerson D.A.
Direct Submission
Submitted (11-Jul-2002) Genome Sciences, University of Washington,
1705 NE Pacific, Seattle, WA 98195, USA
To cite this work please use: NIEHS-SNPs, Environmental Genome
Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA
(URL: http://egp.gs.washington.edu).
Location/Qualifiers
1..5917
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1..274
/rpt_family="Alu"
/rpt_type="dispersed"
170
/frequency="0.09"
/replace="a"
249
/frequency="0.03"
/replace="t"
257
/frequency="0.08"
/replace="a"
267
/frequency="0.04"
/replace="a"
275..402
/rpt_family="Alu"
/rpt_type="dispersed"
562..839
/region not scanned for variation"
593..760
/rpt_family="Alu"
/rpt_type="dispersed"
625
/frequency="0.05"
/replace="a"
782..1108
/rpt_family="Alu"
/rpt_type="dispersed"
977
/frequency="0.05"
/replace="t"
1047
/frequency="0.01"
/replace="a"
1131..1331
/rpt_family="MIR"
/rpt_type="dispersed"
1166
/frequency="0.04"
/replace="t"
1307..1827
/region not scanned for variation"
1410..1524
/rpt_family="MIR"
/rpt_type="dispersed"
1851..5390
/gene="SEI1"
/gene="SEI1"
/product="SCDK4-binding protein p34SEI1"
1887
/gene="SEI1"
/frequency="0.08"
/replace="c"

```


variation 2038 /gene="SEI1" /frequency="0.44" /replace="a" 2086 /gene="SEI1" /frequency="0.01" /replace="g" 2113 /gene="SEI1" /frequency="0.01" /replace="c" 2343 /gene="SEI1" /note="Region not scanned for variation" 2610.3252 /rpt_family="I2" /rpt_type-dispersed 2810.3958 /gene="SEI1" /note="Region not scanned for variation" 2847 /gene="SEI1" /frequency="0.09" /replace="g" 3178 /gene="SEI1" /frequency="0.07" /replace="g" 3250.3755 /gene="SEI1" /note="Region not scanned for variation" 3425.3490 /rpt_family="Mariner" /rpt_type-dispersed 3529.3672 /rpt_family="Alu" /rpt_type-dispersed 3683.3756 /rpt_family="Alu" /rpt_type-dispersed 3822.4112 /rpt_family="Alu" /rpt_type-dispersed 4036.4448 /gene="SEI1" /note="Region not scanned for variation" 4144.4250 /rpt_family="MIR" /rpt_type-dispersed 4303.5013 /gene="SEI1" /codon_start=1 /product="SCDK4-binding protein p34SEI1" /db_xref="GI:21914610" /translation="MISGLKRRKEEREKPLAVDSWMLDPGHAAYAAVAPAYASS LFDLSVAKIHNSLOOSEPDRLHYLVYNTLRTQASMAPPAALPPVSPGAAVADN LLAASDAISASMASLIEDUSHIGLSOAPQPLADEGPPGHSIGGAAPSGAIDLILGP ATGCLDLDGLGLEDIDTSMYDNEMLAPABEGIKPGEDEPGKEAPELDEAILEDYL MDVLVGTQALERPPGPR"

variation 4393 /gene="SEI1" /frequency="0.44" /replace="a" 4788 /gene="SEI1" /frequency="0.08" /replace="c" 4812 /gene="SEI1" /frequency="0.31" /replace="c"

variation 4961 /gene="SEI1" /frequency="0.31" /replace="c"

repeat_region 5391.5553 /rpt_family="MIR" /rpt_type-dispersed 5398 /frequency="0.01" /replace="c" 5449 /frequency="0.01" /replace="a" 5494 /frequency="0.09" /replace="a" 5576.5699 /rpt_family="Alu" /rpt_type-dispersed 5700.5917 /rpt_family="Alu" /rpt_type-dispersed

BASE COUNT 1390 a 1571 c 1536 g 1420 t

ORIGIN

Query Match 87.9%; Score 1048; DB 9; Length 5917; Best Local Similarity 100.0%; Pred. No. 1.8e-221; Matches 1048; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

138 AGATGCTGAGCAAGGGTGTGAAGCGGAACGGGAGAGAGAGAGAGAGAACCTCTG 197
|||||
4301 AGATGCTGAGCAAGGGTGTGAAGCGGAACGGGAGAGAGAGAGAGAGAACCTCTG 4360
|||||
198 CAGTGCAGCTCTGGGGGTAGATCTGGCCACGAGGGGGGACAGGACCCCGGCGG 257
|||||
4361 CAGTGCAGCTCTGGGGGTAGATCTGGCCACGAGGGGGGACAGGACCCCGGCGG 4420
|||||
258 TGGCCTGAGCTCCCTCTTTGACCTCTGAGTGTCTCAAGCTCCACACAGCTGACGAGA 317
|||||
4421 TGGCCTGAGCTCCCTCTTTGACCTCTGAGTGTCTCAAGCTCCACACAGCTGACGAGA 4480
|||||
318 GTGAGCCGAGCACTCGGACCTGCTGTGCTGTGGAACACTCTCGGCGCATCCAGGCT 377
|||||
4481 GTGAGCCGAGCACTCGGACCTGCTGTGCTGTGGAACACTCTCGGCGCATCCAGGCT 4540
|||||
378 CCATGGCACCCGGGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 437
|||||
4541 CCATGGCACCCGGGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 4600
|||||
438 CTGACAACTTACTGGCAAGCTGAGCGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 497
|||||
4601 CTGACAACTTACTGGCAAGCTGAGCGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 4660
|||||
498 AGGACCTGAGCAATGAGGGGCTGAGTACAGTCCCAACCTTGGCAGAGAGGGGC 557
|||||
4661 AGGACCTGAGCAATGAGGGGCTGAGTACAGTCCCAACCTTGGCAGAGAGGGGC 4720
|||||
558 CACGAGCCGCTGAGCATGGGGGAGCAGCGCCAGACCTGCTGCTGCTGCTGCTGCTG 617
|||||
4721 CACGAGCCGCTGAGCATGGGGGAGCAGCGCCAGACCTGCTGCTGCTGCTGCTGCTG 4780
|||||
618 CAGCCTGCTGCTTACTGAGCATGGGCTTGGGGCTTGGAGATATTGACACT 677
|||||
4781 CAGCCTGCTGCTTACTGAGCATGGGCTTGGGGCTTGGAGATATTGACACT 4840
|||||
678 CTATGATGACAAATGAACTTTGGGACACCTCTTGGAGGCTTCAAAACAGGCGCTGAGG 737
|||||
4841 CTATGATGACAAATGAACTTTGGGACACCTCTTGGAGGCTTCAAAACAGGCGCTGAGG 4900
|||||
738 ATGGCCGCGCAAGAGAGAGCTCCGAGCTGAGAGAGGCGGAATTTGACTTACTCTATGG 797
|||||
4901 ATGGCCGCGCAAGAGAGAGCTCCGAGCTGAGAGAGGCGGAATTTGACTTACTCTATGG 4960
|||||
798 ATGTGCTGTGTGGCAGACAGGCACTGAGAGGACCGCGGGGCCAGGGCGCTGAGCCCTCG 857
|||||
4961 ATGTGCTGTGTGGCAGACAGGCACTGAGAGGACCGCGGGGCCAGGGCGCTGAGCCCTCG 5020
|||||

| | | | |
|----|------|---|------|
| Oy | 858 | GGCGGATGGTGGTGGTATCGAATGAGCGCGTGGAGCAACGTCTCTGAA | 917 |
| Db | 5021 | TGCTGGATGGTGTCTGGTATCTGAACTGAGCGCTGCTGGTGACCAACTGTCTCGAA | 5080 |
| Oy | 918 | AAGACACAGCTGGGCTCCCTAGTACAGAGAACAGGGCTTGGGCCACTTGGAGAGACGA | 977 |
| Db | 5081 | AAGACACAGCTGGGCTCCCTAGTACAGAGAACAGGGCTTGGGCCACTTGGAGAGACGA | 5140 |
| Oy | 978 | ATCTAGTCCCTGGGCAACTTCACATCCGTCCTCTGTCTCAGAGGCTGGCAGGGAGAGCCTG | 1037 |
| Db | 5141 | ATCTAGTCCCTGGGCAACTTCACATCCGTCCTCTGTCTCAGAGGCTGGCAGGGAGAGCCTG | 5200 |
| Oy | 1038 | GAATTTACCCCTAGTAGATGAAATGACAGAGGCTGTGGTGGGACTGAATTTCCCTGGCCCTGG | 1097 |
| Db | 5201 | GAATTTACCCCTAGTAGATGAAATGACAGAGGCTGTGGTGGGACTGAATTTCCCTGGCCCTGG | 5260 |
| Oy | 1098 | GGTCAATAGCTTGGGCTGTCTCTCTCTCTGATACGGGAGAGACCCCAATCAGATTTTCAA | 1157 |
| Db | 5261 | GGTCAATAGCTTGGGCTGTCTCTCTCTCTGATACGGGAGAGACCCCAATCAGATTTTCAA | 5320 |
| Oy | 1158 | ATTAAGCCACTCCTGGGAAATCTCCAAA | 1185 |
| Db | 5321 | ATTAAGCCACTCCTGGGAAATCTCCAAA | 5348 |

| RESULT | LOCUS | DEFINITION |
|--------|------------|---|
| 8 | AC010271/c | Homo sapiens chromosome 19 clone CTC-492K19, complete sequence. |
| | AC010271 | 160643 bp DNA linear PRI 13-JUL-2002 |

| | |
|----------|----------------------|
| SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens |

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center
TITLE Direct Submission

REFERENCE 2 (bases 1 to 160643)

TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint

REFERENCE 3 (bases 1 to 160643)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center

Submitted (29-AUG-2000) DOE Joint Genome Institute, 2800 Mitchell

REFERENCE
4 (bases 1 to 160643)
DOE Joint Genome Institute and Stanford Human Genome Center

JOURNAL
Submitted (03-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598 USA

REFERENCE
DOE Joint Genome Institute and Stanford Human Genome Center
AUTHORS
(pages 1 to 100043)

JOURNAL Submitted (13-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Draft Sequence Produced by DOE Joint Genome Institute

FINISHING completed at Stanford Human Genome Center
www-shgc.stanford.edu

FEATURES
Estimated Total Number of Errors is 0.1.
Location/Qualifiers

source

```
/organism="Homo sapiens"  
/mol_type="genomic DNA"  
/db_xref="taxon:9606"  
/chromosome="19"
```

| BASE COUNT | 39112 | a | 42447 | c | 40532 | g | 38552 | t |
|------------|---------------------|---|-------|---|-------|---|-------|---|
| ORIGIN | /clone="CTC-492K19" | | | | | | | |

| | | | | |
|-----------------------|--------------|---------------------|---------------|----------------|
| Query Match | 87.8%; | Score 1046.4; | DB 9; | Length 160643; |
| Best Local Similarity | 99.9%; | Pred. No. 2.9e-221; | | |
| Matches 1047; | Conservative | 0; | Mismatches 1; | Indels 0; |
| | | | | Gaps 0; |

QY 138 AGATGCTGAGCAAGGGGTCTGAAAGCGGAACCGGGAGGAGGAGGAGAGAAAGGAACTCTGG 197

108

Db 118861 CAGTGCACCTGGTGGCTAGATCCTGGCCACACAGCGGTGGCACAGGCAACCCCGCGG 118862

258 TGGCCTCTAGCTCCCTCTTTTGACTCTCAGTGCTCAAGCTCCACCACAGCCTGCAGCAGA 317

[illegible]

Db 118741 GTGAGCCGGACCTGCGGACCTGCTGCTGTCGACACACTCTGCGGGGCGATCCAGGGCT 1188

378 CCATGGACCCGGGCTGCCCTGCCACCTGTGCCCTAGCCACCTGCAGCCCCAGTTGG 437

UN
T10081 CCAAGGACCCCGGGCAGCCCAGCACCTGAGGCACCCAGCCCCAACATGTGG TTAG

Db 118621 CTTGACAACTTACTGCGCAAGCTCGGACGCTGCCCCCTTTCAGGCCCTTCATGCCCCAAGCCTCTTGG 11895

498 AGGACCTCAGCCACATTGAGGGCTGAGTCAGGCTCCCAACCTTGGCAGACGAGGGC 557

DB 110501 AGGACCTCAGCCACATTCAGAGGCCAGATCAGGCTCCCAACCCCTTGGCAGACGAGGGGC 118

Db 118501 CACCAGGCGCTATCGGGGAGCTAGCGCCACGCTGGTGTCTTGGACCTGCTGGGGCC 118511

618 CAGCCACTGGCTGTCTACTGAGCATGGGCTTGAGGGCCGTGTTGAGGATATTGACACCT 677

Db 11841 CAGCCAC16GCTGTCTACTGACGATGGGCTTGAGGGCTCTTGGAGGATATTGACACCT 1183

[illegible]

738 ATGGGCGGGCAAGGAGGAGCTCCGGAGCTGGACGAGGCCGAATTGACCTACCTCATGG 797

Db 118321 ATGGGGCCAGGAGGAACTCCGGAGCTGGACGAGGCCGAAATTGGACTACCTCATGG 118322

[illegible]

858 TCGTGGAATGGTTGCTCTGATCTGAACTGAGCCTGCTGGCTGACCACTGTCTCGAA 917

Db 118201 TGCTGGAATGCTTGTCTGGTATCTGAAC TGAAGCCCTGCTGGCTGGACCAACTGTCTCGAA 118202

[illegible]

978 ATCTAGTCTGGGCACTTCAATCCGTCCTCTGTTCTAGGGCTGGCAGGGGAGCCCTG 10373

Db 118081 ATCTAGTCCCTGGCACTTCACATCCGTCCTCTGTCTCAGGGCTGGCAGGGGAGCCCTG 1180

1038 GGAATACCCCTAGTGAATGATACAGGGTCTGGTGGGACAGAAATCCCTGGCCCTGG 1097

1098 GGTCAATAGCTTGGGCTGTTCCCTTCTCTGATACGGGAAGAGACCCCAATCAGATTTTTCAA 1157

| LOCUS | AX472956 | 5746 bp | DNA | linear | PAT 09-AUG-2002 |
|-----------------------|--|---|-----|--------|-----------------|
| DEFINITION | Sequence 5 from Patent WO0218576. | | | | |
| ACCESSION | AX472956 | | | | |
| VERSION | AX472956.1 | GI:22207743 | | | |
| KEYWORDS | | | | | |
| SOURCE | Homo sapiens (human) | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| REFERENCE | 1 | | | | |
| AUTHORS | Chen, S. Y., Macina, R. A., Sun, Y. and Recipon, H. | | | | |
| JOURNAL | Compositions and methods relating to lung specific genes Patent: WO 0218576-A 5 07-MAR-2002; Diadexus, Inc. (US) | | | | |
| FEATURES | location/Qualifiers | | | | |
| Source | 1..5746 | | | | |
| BASE COUNT | 1141 a 1660 c 1980 g 965 t | | | | |
| ORIGIN | | | | | |
| Query Match | 65.3%; Score 777.8; DB 6; Length 5746; | | | | |
| Best Local Similarity | 99.7%; Pred. No. 1.1e-161; | | | | |
| Matches | 779; Conservative 0; Mismatches 2; Indels 0; Gaps 0; | | | | |
| Y | 26 | CGCGACCCAGGCGGGTTTCGGAGGACAGCCAAACGATAGCGCGCGCTTTC | 85 | | |
| Db | 1 | CGCGACCCAGGCGCGGGTTTCGGAGGACAGCCAAACGATAGCGCGCGCTTTC | 60 | | |
| Y | 86 | CTGATTGGTTGGGTGGCTACCTTTCGTTTGTATGGCCGCTAGTAGCAATAGCTG | 145 | | |
| Db | 61 | CTGATTGGTTGGGTGGCTACCTTTCGTTTGTATGGCCGCTAGTAGCAATAGCTG | 120 | | |
| Y | 146 | AGCAAGGTCGGAAGCGGAAACGGAGAGAGAGAGAGAAAGAACTTGGCAGTCGAC | 205 | | |
| Db | 121 | AGCAAGGTCGGAAGCGGAAACGGAGAGAGAGAGAGAAAGAACTTGGCAGTCGAC | 180 | | |
| Y | 206 | TCTGTGTGGCTAGATCTCTGGCCACGACGCGTGGACAGGACCCCGGCTGTGCTCT | 265 | | |
| Db | 181 | TCTGTGTGGCTAGATCTCTGGCCACGACGCGTGGACAGGACCCCGGCTGTGCTCT | 240 | | |
| Y | 266 | AGCTCCCTCTTGAACCTCTCAAGTGTCAAGTCTCACACACAGCTGACAGAGTAGCCG | 325 | | |
| Db | 241 | AGCTCCCTCTTGAACCTCTCAAGTGTCAAGTCTCACACACAGCTGACAGAGTAGCCG | 300 | | |
| Y | 326 | GACCTGGCGACCTGCTGCTGATGCTGAACACTTGGCGGCATCGACGCTTCATAGGCA | 385 | | |
| Db | 301 | GACCTGGCGACCTGCTGCTGATGCTGAACACTTGGCGGCATCGACGCTTCATAGGCA | 360 | | |
| Y | 386 | CCGCGGCGCTGCGCACCTGTGTGCTTGAAGCCACCTGACAGCCCGCAGTGTGCTGCAAC | 445 | | |
| Db | 361 | CCGCGGCGCTGCGCACCTGTGTGCTTGAAGCCACCTGACAGCCCGCAGTGTGCTGCAAC | 420 | | |
| Y | 446 | TTACTGGCAACTGGAGAGCTGCGCTTTCAGAGCTTCATAGGCGCAAGCTCTTGAAGAGCTC | 505 | | |
| Db | 421 | TTACTGGCAACTGGAGAGCTGCGCTTTCAGAGCTTCATAGGCGCAAGCTCTTGAAGAGCTC | 480 | | |
| Y | 506 | AGCACAATTGAGGGCTGAGTCAAGGCTCCCAACCTTGGCGAGAGAGGGGCTCACAGGC | 565 | | |
| Db | 481 | AGCACAATTGAGGGCTGAGTCAAGGCTCCCAACCTTGGCGAGAGAGGGGCTCACAGGC | 540 | | |
| Y | 566 | CGTAGCATCGGGGAGACGCGCCACCTGTGGGTGCTTGGACCTGTGGGCCACCACT | 625 | | |

| | | | | | |
|-----------------------|--|---|-----------|-------------|-----------------|
| Db | 541 | CGTAGCATCGGGGAGGACGAGCCGCCAGCCTGGGTGGCTTGGACCTGCTGGCCCACT | 600 | | |
| QY | 626 | GGCTGTCTACTGACGATGGCCTTGAGGGCCTGTTTGAGATATTGACACCTATGTAT | 685 | | |
| Db | 601 | GGCTGTCTACTGACGATGGCCTTGAGGGCCTGTTTGAGATATTGACACCTATGTAT | 660 | | |
| QY | 686 | GACATGAACCTTTGGGGACCAAGCCCTGATAGGGCCCAACAAGGCCCGAGATGGGCG | 745 | | |
| Db | 661 | GACATGAACCTTTGGGGACCAAGCCCTCTTAGGGCCTCAACAAGGCCCTGAGAGATGGGCG | 720 | | |
| QY | 746 | GGCAAGAGGAAGCTCCGAGCTGTGACGAGGCCGGAATTGACTTACCTCATGGATGGTG | 805 | | |
| Db | 721 | GGCAAGAGGAAGCTCCGAGCTGTGACGAGGCCGGAATTGACTTACCTCATGGATGGTG | 780 | | |
| QY | 806 | G | 806 | | |
| Db | 781 | G | 781 | | |
| RESULT 10 | HSA323696 | 978 bp | DNA | linear | PRI 18-JUL-2001 |
| LOCUS | HSA323696 | | | | |
| DEFINITION | Homo sapiens genomic sequence surrounding NotI site, clone | | | | |
| ACCESSION | NB6-544S. | | | | |
| VERSION | AJ323696 | | | | |
| KEYWORDS | AJ323696.1 GI:15868075 | | | | |
| SOURCE | Homo sapiens (human) | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 978) | | | | |
| AUTHORS | Kusenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Krasha,S.M., Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V., Levitsky,V.G., Kolchanov,N.A., Protodopov,A.I., Kashba,V.I., Kiselev,I.L., Wasserman,W., Wahlestedt,C. and Zabarovskiy,E.R. | | | | |
| TITLE | NotI flanking sequences: a tool for gene discovery and verification of the human genome | | | | |
| JOURNAL | Nucleic Acids Res. | 30 | (14), | 3163-3170 | (2002) |
| MEDLINE | 22131767 | | | | |
| PUBMED | 12136098 | | | | |
| REFERENCE | 2 (bases 1 to 978) | | | | |
| AUTHORS | Zabarovskiy,E.R. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (16-May-2001) Microbiology and Tumorbiology Centre, Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77, Sweden | | | | |
| FEATURES | | | | | |
| source | Location/Qualifiers | | | | |
| | 1..978 | | | | |
| | /organism="Homo sapiens" | | | | |
| | /mol_type="genomic DNA" | | | | |
| | /db_xref="taxon:9606" | | | | |
| | /clone="NB6-544S" | | | | |
| BASE COUNT | 188 a | 300 c | 299 g | 188 t | 3 others |
| ORIGIN | | | | | |
| Query Match | 61.9%; | Score 738; | DB 9; | Length 978; | |
| Best Local Similarity | 96.8%; | Pred. No. 9e-153; | | | |
| Matches 763; | Conservative 0; | Mismatches 7; | Indels 2; | Gaps 2; | |
| QY | 200 | GTGCATCTCTGGTGGCTAGATCTGGCCACGACGCGGTGGGCACAGGACCCCGGCGCTG | 259 | | |
| Db | 1 | GTGCATCTCTGGTGGCTAGATCTGGCCACAGCGGTGGGCACAGGACCCCGGCGCTG | 60 | | |
| QY | 260 | GCCCTAGCTCCCTCTTTGACCTCTCAGTGCATGCTCCACCAAGCCTGCACAGAGT | 319 | | |
| Db | 61 | GCCCTAGCTCCCTCTTTGACCTCTCAGTGCATGCTCCACCAAGCCTGCACAGAGT | 120 | | |
| QY | 320 | GAGCGGACCTGGCGGACCTGTGCTGTGTGTAACACTCTGCGGGCGCATCAAGCGTCC | 379 | | |
| Db | 121 | GAGCGGACCTGGCGGACCTGTGCTGTGTGTAACACTCTGCGGGCGCATCAAGCGTCC | 180 | | |

QY 380 ATGGACACCGCGGCTCCCTGCCACTGTGCTTACGCCACCTGACAGCCGCCCAAGTGGCT 439
 DB 181 ATGGACACCGCGGCTCCCTGCCACTGTGCTTACGCCACCTGACAGCCGCCCAAGTGGCT 240
 QY 440 GACAACTTACTGGCAAGCTGGAGCGCTTGCCTTACGCTTCACTGGCCAGCCCTCCCTGGAG 499
 DB 241 GACAACTTACTGGCAAGCTGGAGCGCTTGCCTTACGCTTCACTGGCCAGCCCTCCCTGGAG 300
 QY 500 GACCTGACGACATTTAGAGGCTTGAAGTCAAGGCTCCCAACCCCTTGGCAGACAGAGGGGCA 559
 DB 301 GACCTGACGACATTTAGAGGCTTGAAGTCAAGGCTCCCAACCCCTTGGCAGACAGAGGGGCA 360
 QY 560 CCAGGCGCTTACGATCGGGGAGACAGCGCCAGCTTGGTGGCTTGGAGCTTGGGCGCA 619
 DB 361 CCAGGCGCTTACGATCGGGGAGACAGCGCCAGCTTGGTGGCTTGGAGCTTGGGCGCA 420
 QY 620 GCCACTGGCTTACTTGGACGATGGGCTTGAAGGCTTGGAGGATTTGAGACCTCT 679
 DB 421 GCCACTGGCTTACTTGGACGATGGGCTTGAAGGCTTGGAGGATTTGAGACCTCT 480
 QY 680 ATGTATGACAAATGAATTTGGGACACGCTCTGAGGCGCTCAAAACAGGCGCTGAGAT 739
 DB 481 ATGTATGACAAATGAATTTGGGACACGCTCTGAGGCGCTCAAAACAGGCGCTGAGAT 540
 QY 740 GGGCGGCGGACAGGAGAGCTCGGAGCTGAGACAGAGCGCAATTTGACTACTATGAT 799
 DB 541 GGGCGGCGGACAGGAGAGCTCGGAGCTGAGACAGAGCGCAATTTGACTACTATGAT 600
 QY 800 GTGCTGTGGGACACAGGACATGGAGGAGCGGCGGCGGCGGAGGCGTACCCCTCG 859
 DB 601 GTGCTGTGGGACACAGGACATGGAGGAGCGGCGGCGGCGGAGGCGTACCCCTCG 660
 QY 860 CTGGAATGGTCTTCTGTATCTGAATGAGCTGTGCTGAGACCAACTGCTCGAATA 919
 DB 661 CTGGAATGGTCTTCTGTATCTGAATGAGCTGTGCTGAGACCAACTGCTCGAATA 720
 QY 920 GACACAG-CTGGCTTCCCTAGTACAGAGACA-GGGCTTGGGCGCACTTTGA 969
 DB 721 GACACAGCTGCTTCCCTAGTACAGAGAGAGGGGCTTGGGCGCACTTTGA 772

RESULT 11
 AF366401 1247 bp mRNA linear ROD 13-MAY-2001
 LOCUS Mus musculus TRIP-Brl mRNA, complete cds.
 DEFINITION AF366401
 ACCESSION AF366401 GI:14029831
 VERSION AF366401.1
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Hsu,S.I., Yang,C.M., Sim,K.G., Heutschel,D.M., O'Leary,E. and
 Bonventre,J.V.
 TRIP-Brl: a novel family of PHD zinc finger- and
 bromodomain-interacting proteins that regulate the transcriptional
 activity of E2F-1/DP-1
 EMBO J. 20 (9), 2273-2285 (2001)
 JOURNAL MEDLINE
 PUBLISHED 21231173
 11331592
 2 (bases 1 to 1247)
 Hsu,S.I. and Bonventre,J.V.
 REFERENCE
 TITLE Direct Submission
 JOURNAL Submitted (29-MAR-2001) Dept. of Medicine, National University
 Hospital, 5 Lower Kent Ridge Road 119074, Singapore
 FEATURES
 source
 1..1247
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="dbEST:W82777"
 /db_xref="taxon:10090"
 173..883

BASE COUNT 276 a 352 c 372 g 247 t
 ORIGIN
 Query Match 56.7%; Score 675.6; DB 10; Length 1247;
 Best Local Similarity 76.4%; Pred. No. 5,6e-139;
 Matches 931; Conservative 0; Mismatches 254; Indels 33; Gaps 7;
 /note="mTRIP-Brl; transcriptional regulator interacting
 with the PHD-bromodomain; similar to Homo sapiens TRIP-Brl
 (SEI-1)"
 /codon_start=1
 /product="TRIP-Brl"
 /protein_id="AAK52830.1"
 /db_xref="GI:14029832"
 /translation="MISGLKRRREEETMEALSYDSCLDPSHPVAQTPPTVAS
 LFDLSVKKLHSLROSEPDRLILVYVNTLRIOAMSEPAVLPEEPAPVADS
 ILASDAGLSASLIEDNLINEDNLNPOADBPGRGRTGTSPLGLDLGP
 ATGCLDLSDELFEIDTSMYDSELMPLASGLRGPENGRFAKEPELDRAEIDL
 MDVLGTOALERPDPGR"
 1 GTGAGCGGAGTCAAGAACTGCTGCGACGACGCGGCGGCTTCCGAGACAGCCAA 60
 DB 31 GTGCGCGGAGTCAAGAACTGCGCTACGCTCCAGGTGCTGCTGAGACAGAAAG 90
 QY 61 CAAGGATGCTGCGCGCGCTTCTGATTTGTTGGTGGCTACTCTTCTGTTGA 120
 DB 91 CTAGGATGCTACGCGACGCGCTTCTGATTTGTTGGTGGCTACTCTTCTGTTGA 150
 QY 121 TTGGCGCGTGA---GCAAGATGCTGAGCAAGGCTCTGAAGCGGAAACGGAGAGGA 177
 DB 151 TTGGCGCGTGAAGTCTGAGAGTCTGAGCAAGGCTCTGAAGCGGAAACGGAGAGGA 210
 QY 178 GGAGGAGAGAACTCTGGAGTCACTGCTGAGTCAAGTCTGAGTCAAGTCTGAGT 237
 DB 211 GGAGGAGAGAACTCTGGAGTCACTGCTGAGTCAAGTCTGAGTCAAGTCTGAGT 270
 QY 238 GGCACAGGACCCCGCGCGCTTCTGATTTGTTGGTGGCTACTCTTCTGTTGA 297
 DB 271 GGCACAGGACCCCGCGCGCTTCTGATTTGTTGGTGGCTACTCTTCTGTTGA 330
 QY 298 CCACCAAGGCTGAGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 357
 DB 331 CCACCAAGGCTGAGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 390
 QY 358 TCTGGCGGATCAAGAGGCTCCATGAGGACCCCGCGCTTCTGATTTGTTGGTGG 417
 DB 391 ACTGAGGCGATCAAGAGGCTCCATGAGGACCCCGCGCTTCTGATTTGTTGGTGG 450
 QY 418 ACTGAGGCGGATCAAGAGGCTCCATGAGGACCCCGCGCTTCTGATTTGTTGGTGG 477
 DB 451 GGCACAGGACCCCGCGCGCTTCTGATTTGTTGGTGGCTACTCTTCTGTTGA 510
 QY 478 CTGCAAGGACGCTCTGAGAGGCTCAAGCAATGAGGCGTCAAGTCAAGTCAAGT 537
 DB 511 CTGCAAGGACGCTCTGAGAGGCTCAAGCAATGAGGCGTCAAGTCAAGTCAAGT 570
 QY 538 ACCCTTGGAGAGAGGCGCCACAGGCGCTTCTGATTTGTTGGTGGCTCAAGTCAAG 597
 DB 571 ACCCTTGGAGAGAGGCGCCACAGGCGCTTCTGATTTGTTGGTGGCTCAAGTCAAG 630
 QY 598 TGCCTTGGAGAGGCGCCACAGGCGCTTCTGATTTGTTGGTGGCTCAAGTCAAG 657
 DB 631 TGCCTTGGAGAGGCGCCACAGGCGCTTCTGATTTGTTGGTGGCTCAAGTCAAG 690
 QY 658 GTTGGAGATTTGAGAGGCTTCTGATTTGTTGGTGGCTCAAGTCAAGTCAAGT 717
 DB 691 GTTGGAGATTTGAGAGGCTTCTGATTTGTTGGTGGCTCAAGTCAAGTCAAGT 750
 QY 718 CCTCAAGAGGCGCTTCTGAGAGGCGCCACAGGCGCTTCTGATTTGTTGGTGG 777
 DB 751 TCTCAAGAGGCGCTTCTGAGAGGCGCCACAGGCGCTTCTGATTTGTTGGTGG 810
 QY 778 CGAATTGAGATTTCTGATTTGTTGGTGGCGACAGGCGTCAAGTCAAGTCAAGT 837

| | | | |
|----|------|--|------|
| Db | 811 | GGAGTGTGACACACTCAATGAGATGACAGTGTAGCGACACACAGGCACTGGAAAGGCCACCAGG | 870 |
| QY | 838 | GCCAGGGGGCTGAGACCCCTCG--TGCTGGAAATGGTTGTGTGATCTGAACCTGAGCCTGCT | 895 |
| Db | 871 | GCCGTGGGGGCTGACCCCTTAGGACCCAGGGAATAATGACTTGGATCTGGTTGAGGCTGTG | 930 |
| QY | 896 | GGCTGAGACCACTGCTGCTCGAAAGAGACACAGCCTGGCTCCCTAGTACAGAGAACAGGGCT | 955 |
| Db | 931 | GGTGACACCAACCCCACTTTGACACGACACAGGTGGCGTCCC--CTCACAGAGAGAGACC | 987 |
| QY | 956 | TGGGCCCACTTTGGAGAGACAGAAATCTAGTCTCTGGGCAACTTCACATCCGTCCTCTGTCT | 1015 |
| Db | 988 | AGGGTCACTTTGGAGAGACTGTACTAGATGCTCGGG--AACTTCACATTCATCCACCTGTCT | 1048 |
| QY | 1016 | CAGGGCTGGCAGGGGGAGCCTGGAAATTCACCCCTAGTATGTGAATGAGAGGGTC----- | 1065 |
| Db | 1047 | ---TGACCAGTGAAGAGATGTGAATATGAGCCCGAGTGCTTGAAATGACAGAGGCCCATGTA | 1103 |
| QY | 1070 | -----TGGTGGGGCACTAATTTCCCTGGGCCCTGGGGGTATAGACTTGGGGCTGTTCC-- | 1118 |
| Db | 1104 | CAGTAATAGTGTTTGTGGGCGCAGGCTGGGGTTTGAATCGGATATATAGTTTGGGGCTATCCAT | 1163 |
| QY | 1119 | ----TTCTCTGATACGGGAGAGACCCCAATCAGATTTTTCAAATTAAGCCAGTCTGTGG | 1174 |
| Db | 1164 | CCCATTTATCCCACTGTGTGGAGAGAGACCTCATTAATTTTGAATAATTAACACAGCTTTGG | 1223 |
| QY | 1175 | GAATCTCAAAAAAAAAA | 1192 |
| Db | 1224 | GAATCAAAAAAAAAA | 1241 |

| | |
|------------|-------------------------------------|
| RESULT | 12 |
| AX537459 | |
| LOCUS | AX537459 |
| DEFINITION | Sequence 30 from Patent WO02070709. |
| ACCESSION | AX537459 |
| VERSION | AX537459.1 GI:25269253 |
| KEYWORDS | |
| SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens |

| REFERENCE | AUTHORS | TITLE | JOURNAL | FEATURES | SOURCE | BASE COUNT | ORIGIN |
|-----------|--|---|---------------------------------------|----------|--------|------------|--------|
| 1 | Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo. | | | | | | |
| | Lai, P. G., Baughn, M. R., Yao, M. G., Walla, N. K., Elliot, V. S., Xu, Y., Honchell, C. D., Yue, H., Ding, T., Gietzen, K. J., Ison, C. H., Lu, D. A., Hacheila, A. J., Gandhi, A. R., Thangavelu, K., Sanjanvala, M. M., Tang, Y. T., Ramkumar, J., Griffin, J. A., Svarnaker, A., Azimail, Y., Saperstein, S. K., Burford, N., Lee, E. A., Lu, Y., Tran, U. K. and Marquis, J. P. | Methods for disease detection and treatment | Patent: WO 02070709-A 30 12-SEP-2002; | | | | |
| | Incyte Genomics, Inc. (US) | | | | | | |
| | Location/Qualifiers | | | | | | |
| | 1..5869 | | | | | | |
| | /organism="Homo sapiens" | | | | | | |
| | /mol_type="genomic DNA" | | | | | | |
| | /db_xref="taxon:9606" | | | | | | |
| | /note="Incyte ID No: 290316CB1" | | | | | | |
| | 1191 a 1682 c 1995 g 1001 t | | | | | | |

| | | | | | | |
|----|--|-----------------|---------------------|-----------|--------------|--|
| | Query Match | 56.0%, | Score 667.4; | DB 6; | Length 5869; | |
| | Best Local Similarity | 99.9%; | Pred. No. 3.1e-137; | | | |
| | Matches 668; | Conservative 0; | Mismatches 1; | Indels 0; | Gaps 0; | |
| QY | 138 AGAGTCGTACCAAGGCTCTGAAGCGAAGCGGAGAGAGAGAGAGAACAACCTTGG | 197 | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| Db | 237 AGAGTCGTACCAAGCGCTCTGAAGCGAAGCGGAGAGAGAGAGAGAACAACCTTGG | 296 | | | | |
| | | | | | | |
| | | | | | | |
| QY | 198 CAGTCGACTCCTTGTTGCTTAGATCCTGGCCACGCAGCGGTGGCACAGGCCACCCTCCGGCG | 257 | | | | |
| | | | | | | |
| | | | | | | |
| Db | 297 CAGTCGACTCCTTGTTGCTTAGATCCTGGCCACACAGCGGTGGCACAGGCCACCCTCCGGCG | 356 | | | | |
| | | | | | | |
| | | | | | | |

| | | | |
|----|-----|---|-----|
| QY | 258 | TGGCCTCAGACCTCCCTCTTTTACGCMCAGTGCCTCAAGCTCCACACACGCTGGAGAGA | 317 |
| Db | 357 | TGGCCTCAGTCCCTCTTTGACCTCTCAGTGCCTCAAGCTCCACACACGCTGGAGAGA | 416 |
| QY | 318 | GTGAGCCGAGCTCGGGACCTGGTGTGTCGTGCTGGAACACTCTCGGGCGACTCCAGGCGT | 377 |
| Db | 417 | GTGAGCCGAGCTCGGGACCTGGTGTGTCGTGCTGGAACACTCTCGGGCGACTCCAGGCGT | 476 |
| QY | 378 | CCATGGACCCGGGCTGCCTTGCACCTGTGCTAGCCACCTGACAGCCCGCCAGTGTGG | 437 |
| Db | 477 | CCATGGACCCGGGCTGCCTTGCACCTGTGCTAGCCACCTGACAGCCCGCCAGTGTGG | 536 |
| QY | 438 | CTGACAACTTACTGGCAAGCTCGGACGCTGCCCTTTACGCTCCATGGCCACGCTCCTGG | 497 |
| Db | 537 | CTGACAACTTACTGGCAAGCTCGGACGCTGCCCTTTACGCTCCATGGCCACGCTCCTGG | 596 |
| QY | 498 | AGGACCTCAGGCACATTGAGGGCGCTGAGTCAAGGCTCCCAACCCCTTGGCAGACGAGGGC | 557 |
| Db | 597 | AGGACCTCAGGCACATTGAGGGCGCTGAGTCAAGGCTCCCAACCCCTTGGCAGACGAGGGC | 656 |
| QY | 558 | CACCAAGCGCCTAGCATCGGGGAGACAGCGCCACACTTGGGTCCTTGGACCTGTGGGCC | 617 |
| Db | 657 | CACCAAGCGCCTAGCATCGGGGAGACAGCGCCACACTTGGGTCCTTGGACCTGTGGGCC | 716 |
| QY | 618 | CAGCCACTGGCTGCTACTGTGAGCATGAGGCTTGAAGGCGCTGTTGAGGATATTGACACT | 677 |
| Db | 717 | CAGCCACTGGCTGCTACTGTGAGCATGAGGCTTGAAGGCGCTGTTGAGGATATTGACACT | 776 |
| QY | 678 | CTATGTATGACAAATGAACTTTGGGCAACCAGCCTTGAGGGCCTCAAAACAGGCCCTGAGG | 737 |
| Db | 777 | CTATGTATGACAAATGAACTTTGGGCAACCAGCCTTGAGGGCCTCAAAACAGGCCCTGAGG | 836 |
| QY | 738 | ATGGGCGCGGGAAGAGAAAGCTCCGAGAGCTGGAGACAGAGCGGAATTGGACTACTCATGG | 797 |
| Db | 837 | ATGGGCGCGGGAAGAGAAAGCTCCGAGAGCTGGAGACAGAGCGGAATTGGACTACTCATGG | 896 |
| QY | 798 | ATGTGCTGG 806 | |
| Db | 897 | ATGTGCTGG 905 | |

| | | | | | |
|------------|--|-------------|------|--------|-----------------|
| RESULT | 13 | | | | |
| LOCUS | BC016077 | | | | |
| DEFINITION | Mus musculus RIKEN CDNA 1110032C13 gene. | 1165 bp | mRNA | linear | ROD 16-APR-2003 |
| ACCESSION | BC016077 | | | | |
| VERSION | BC016077.1 | GI:16359221 | | | |
| KEYWORDS | MGC. | | | | |
| SOURCE | Mus musculus (house mouse) | | | | |
| ORGANISM | Mus musculus | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. | | | | |
| AUTHORS | 1 (bases 1 to 1165) Straussberg,R.L., Feingold,E.A., Grouse,L.H., Dere,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefter,C.F., Hsieh,N.K., Hopkins,R.E., Jordan,H., Moore,T., Max,S.I., Wang,J., Stahl,F., Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stoplecko,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., | | | | |

Schecter, I. E., Brownstein, M. J., Usain, T. B., Tosnyuk, S., Caciucol, P., Prange, C., Raha, S. S., Loquellano, N. A., Peters, G. J., Abramson, R. D., Mullany, S. J., Bosak, A. A., McEwan, P. J., McKernan, K. J., Malek, J. A., Gunnarsson, P. H., Richards, S., Wolter, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hulyk, S. W., Villalón, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A., Fahy, J. J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A. C., Shevchenko, Y., Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, E. D., Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butlerfield, I. S., Krzywinski, M. I., Skalska, U., Smalls, D. E., Scherch, A., Schein, J. E., Jones, S. J., and Maitra, M. A.

JOURNAL human and mouse cDNA sequences
 MEDLINE Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 PUBMED 22388257
 12477932
 REFERENCE 2 (bases 1 to 1165)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (22-Oct-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@gscc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeary, Steven
 Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smallish, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

FEATURES
 source Clon distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://lmln1.nih.gov>
 Series: IRK Plate: 38 Row: C Column: 18
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA g1: 9055331.
 Location/Qualifiers
 1..1165

gene
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:27573 IMAGE:4486318"
 /tissue_type="Mammary tumor. Metastatic lymph node-TGF alpha
 model. 10 month old virgin mouse. Taken by biopsy."
 /clone_1lb="NCI CGAP Mam1"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"
 1..1165
 /gene="1110032C13R1X"
 /note="synonyms: Set1, TRIP-Brl"
 /db_xref="LocusID:55942"
 /db_xref="MGI:1913438"
 90..800
 CDS
 /codon_start=1
 /product="p34SRI-1"
 /protein_id="AAH16077.1"
 /db_xref="GI:16359222"
 /translation="MISKILKREEREETMEALSYDSCWLDSPHAVAQTPPTVSS
 LEFLSVKLHSLRQSEPDRLHLYVNTLRIOASMEAPVLPPEPPOPPASVADS
 ILASDAGISASMASLIEDNIEDLNAPQADBPGRIGTISRLGLDLGLP
 ATGCLDGLBELEFEDIDTSMDSLEMLPASEGLKFPENGPAAKEEPPLDAEIDYL
 MDVLVGTQLERPPGPR"

BASE COUNT 262 a 332 c 341 g 230 t
 ORIGIN

Query Match 53 83; Score 640.8; DB 10; Length 1165;
 Best Local Similarity 76.4%; Pred. No. 2.8e-131;
 Matches 869; Conservative 0; Mismatches 242; Indels 33; Gaps 7;

QY 55 AGCAACAAGCAGTGGCGCGCGGTTCTGATGTTGGTGCGTACGCTCTTG 114
 Db 2 AACAACTGAGTGTAGCGAGCGCGCTTCTGATGTTGGTGCGTCTCTCTTG 61

QY 115 TTCTGATGCGCGCTAGTA--GCAAGTCTGAGCAAGGCTGAGCGAAGCGGA 171
 Db 62 TTCTGATGCGCGCGAGTATCTGTAGTGTGAGCAAGGCTGAGCGAAGCGGA 121
 QY 172 GAGGAGAGAGAGAGAGAACTCTGTGACATGCTCTGTGATGCTGCGACCG 231
 Db 122 GAGGAGAGAGAGAGAGAACTCTGTGACATGCTCTGTGATGCTGCGACCG 181
 QY 232 AGCGGAGAGAGAGAGAGAACTCTGTGACATGCTCTGTGATGCTGCGACCG 291
 Db 182 GCGAGTGGGAGAGAGAGAGAACTCTGTGACATGCTCTGTGATGCTGCGACCG 241
 QY 292 CAAGTCCACACAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 351
 Db 242 CAAGTCCACACAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
 QY 352 GAACACTGCGCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 411
 Db 302 GAACACTGCGCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
 QY 412 TAGCCACCTGAG 471
 Db 362 CATCCAGCCCGAG 421
 QY 472 TTGAGCTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
 Db 422 CTGAGCTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
 QY 532 TCCCAACCTGAG 591
 Db 482 CCCCACACCTGAG 541
 QY 592 CCGGCTGCTGAG 651
 Db 542 CCGGCTGCTGAG 601
 QY 652 GGGCTGTTGAGAGATGATGACACTCTATGATGATGATGATGATGATGATG 711
 Db 602 GGGCTGTTGAGAGATGATGACACTCTATGATGATGATGATGATGATGATG 661
 QY 712 TAGAGGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771
 Db 662 TAGAGGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 721
 QY 772 CGAGGCTGCAAG 831
 Db 722 TAGAGGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 781
 QY 832 GCGGAGGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 889
 Db 782 ACCAGGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
 QY 890 CCTGCTGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 949
 Db 842 CCTGCTGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898
 QY 950 AGGCTGCTGCAAG 1009
 Db 899 AGGCTGCTGCAAG 957
 QY 1010 CTGCTGCAAG 1069
 Db 958 CTGCTGCAAG 1014
 QY 1070 TGTGAGGAG 1114
 Db 1015 CAGTGAAG 1074
 QY 1115 TTCC-----TCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1168
 Db 1075 TCCATCCATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1134
 QY 1169 TCCATCCATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1192

| | | | |
|----|-----|---|-----|
| OY | 1 | GTGACCCGAGTCAAGAACTGCGTCGCGACACCGAGCGGGGTTTCCGGAGGACAGCCAA | 60 |
| | | | |
| Db | 21 | GTGACCCGAGTCAAGAACTGCGTCGCGACACCGAGCGGGGTTTCCGGAGGACAGCCAA | 80 |
| OY | 61 | CAAGCGATGCTGCCCCGCCCTTTCCTGATTTGGTTGTGGGTGGGACCTCTTCTGTTCTGA | 120 |
| | | | |
| Db | 81 | CAAGCGATGCTGCCCCGCCCTTTCCTGATTTGGTTGTGGGTGGGACCTCTTCTGTTCTGA | 140 |
| OY | 121 | TTGGCCGCTAGTGAGCAAGANTGCTGAGCAAGGGTCTGAAGCGGAACGGGAGGAGGAGA | 180 |
| | | | |
| Db | 141 | TTGGCCGCTAGTGAGCAAGANTGCTGAGCAAGGGTCTGAAGCGGAACGGGAGGAGGAGA | 200 |
| OY | 181 | GGAGAGAGAACTCTGCGCAGTGCAGTCTCTGTGTGCTAGATCTCTGGCCACGACGGTGGC | 240 |
| | | | |
| Db | 201 | GGAGAGAGAACTCTGCGCAGTGCAGTCTCTGTGTGCTAGATCTCTGGCCACAGCGGTGGC | 260 |
| OY | 241 | ACAGGACACCCCGCGCGTGGGCTCTAGTCCCTCTTGAACCTCTGAGTCTCAAGCTCCA | 300 |
| | | | |
| Db | 261 | ACAGGACACCCCGCGCGTGGGCTCTAGTCCCTCTTGAACCTCTGAGTCTCAAGCTCCA | 320 |
| OY | 301 | CCACAGCCTGCGAGAGTGAAGCAGGACCTGCGGCACTGAGTCTGGTGTGAACACTCT | 360 |
| | | | |
| Db | 321 | CCACAGCCTGCGAGAGTGAAGCAGGACCTGCGGCACTGAGTCTGGTGTGAACACTCT | 380 |
| OY | 361 | GCGGCGCATCCAGGCGCTCCATGGCAACCCGCGGCTGCGCTGCCACCTGTCCTTAAGCCACC | 420 |
| | | | |
| Db | 381 | GCGGCGCATCCAGGCGCTCCATGGCAACCCGCGGCTGCGCTGCCACCTGTCCTTAAGCCACC | 440 |
| OY | 421 | TGCAGCCCCCACTGTGTGCTGACAACTTACTTGGCAAGCTGGAGCCTGCCCCTTACAGCTTC | 480 |
| | | | |
| Db | 441 | TGCAGCCCCCACTGTGTGCTGACAACTTACTTGGCAAGCTGGAGCCTGCCCCTTACAGCTTC | 500 |
| OY | 481 | CATGGCCAGCCCTCTGGAGAGACCTCAGGCCAATTGAGGGCTCGAGTCAAGGATCCCCAACC | 540 |
| | | | |
| Db | 501 | CATGGCCAGCCCTCTGGAGAGACCTCAGGCCAATTGAGGGCTCGAGTCAAGGATCCCCAACC | 560 |
| OY | 541 | CTTGGCAGACGAGGGGCCACCAAGCGCTTAGCATTCGGGGGAGAGACGCCACAGCTTGAGTGC | 600 |
| | | | |
| Db | 561 | CTTGGCAGACGAGGGGCCACCAAGCGCTTAGCATTCGGGGGAGAGACGCCACAGCTTGAGTGC | 620 |
| OY | 601 | CTTGGACCTGCTGGGCCCAAGCCACTGCGCTGTCTACTGACGATGGGCTTGAAGGCGCTGTT | 660 |
| | | | |
| Db | 621 | CTTGGACCTGCTGGGCCCAAGCCACT-GCTTGCTACTGACGAT-GGCTTGAAGGACTGNT | 678 |
| OY | 661 | TGAGGATTTGACACCT 677 | |
| | | | |
| Db | 679 | TGAGGATTTGACCCCTT 695 | |

Search completed: August 25, 2003, 15:05:57
Job time : 4436 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)